SADEV

SWEDISH AGENCY FOR DEVELOPMENT EVALUATION

Strengthening the Results-Orientation in Sweden's Engagement in Multilateral Development Cooperation

An evaluation of UNDP's country level evaluation activities

Erik Hedblom & Viktoria Hildenwall

SADEV REPORT 2008:6

Swedish Agency for Development Evaluation
P. O. Box 1902, SE-651 19 Karlstad, Sweden

SADEV REPORT 2008:6
Strengthening the Results-Orientation in Sweden's Engagement in Multilateral Development Cooperation
An evaluation of UNDP's country level evaluation activities
Author(s): Erik Hedblom and Viktoria Hildenwall

SADEV Reports are available at www.sadev.se
Printed in Karlstad, Sweden 2008

ISSN 1653-9249
ISBN 978-91-85679-13-3

Preface

The Swedish Agency for Development Evaluation (SADEV) is a government-funded institute that conducts and disseminates evaluations of international development cooperation activities. SADEV's overriding objective is to contribute to the achievement of the goals of Swedish development cooperation.

Evaluation activities at SADEV are conducted along two major strands. The first of these involves the organisation of international development cooperation, and focuses on issues such as the management and monitoring of executive organisation, the choice of modalities, donor coordination and the internal efficiency of donor organisations. The second area is concerned with the short- and long-term impact of development assistance on global poverty. Results of SADEV's evaluations are published in series, which are available electronically from SADEV's website, and in hard copy.

In assessing the quality and management of the evaluations of one multilateral organisation – UNDP – this report makes findings and conclusions on aspects requiring particular attention in order to enable Sweden to have increased confidence in relying on evaluations produced by this organisation.

The report takes as its point of departure the Swedish Strategy for Multilateral Development Cooperation. It provides recommendations for a more results-orientated Swedish involvement in multilateral development cooperation.

Lennart Wohlgemuth
Director General

July 2008

Executive summary

This report sets out the findings of an evaluation of UNDP's evaluation activities at the country level. It draws conclusions about: i) general evaluation quality; ii) the conduct and use of country office (CO) evaluations; and iii) efforts undertaken by UNDP for strengthening evaluation capacities in partner countries. Based on these conclusions, recommendations are directed to the Swedish Ministry for Foreign Affairs. These recommendations focus on three areas for action for a more coherent and results-orientated involvement in multilateral development cooperation: i) both UNDP-specific and general measures for improving Sweden's assessment of the performance of multilateral organisations; ii) measures for a more coherent and strategic Swedish positioning on UNDP's Executive Board; and iii) issues to be addressed in Sweden's organisation-specific strategy in relation to UNDP.

Rationale

The Paris Declaration, and in particular its strengthened focus on results, has given evaluation a central role in today's international development cooperation. At the same time, the Paris principles pose increased challenges to evaluation: evaluation should be nationally owned and conducted in collaboration with other donors. As a signatory to the Paris Declaration, Sweden's policy should of course reflect the principles stipulated in the Declaration. Sweden's Strategy for Multilateral Development Cooperation represents an effort to increase the focus on results-orientation in Sweden's engagement in multilateral development cooperation. As stipulated in the Strategy, multilateral performance should be assessed in terms of relevance and effectiveness. The assessment should concentrate on results at the country level and be primarily based on the multilateral organisations' own reporting systems. This approach places a heavy emphasis on follow-up and evaluation in these organisations, and increases the expectations that results will be reported.

Hence, it is necessary to improve knowledge about UNDP's country level evaluation activities and the overall quality of evaluations produced by its country offices. This should reveal valuable insights about areas potentially requiring improvement in order to improve the level of confidence that Sweden can have in relying upon evaluations produced by UNDP. The study will also reveal issues requiring particular attention in Sweden's relationship with UNDP – as a member of its Executive Board and in Sweden's collaboration with UNDP in general – in order that Sweden can stress the importance of results-orientation within the organisation.

To properly determine the extent to which a multilateral organisation's reporting can be relied upon by Sweden in assessing the organisation's performance, all of the organisation's reporting should be included within the scope of the study. However, this evaluation explicitly focuses on evaluations produced by UNDP at the country level, for a number of reasons. First, it is at the country level that actual development results can be measured. Second, these evaluations constitute important building blocks for higher level evaluations, such as policy or strategic evaluations, which have a higher profile and have more impact, both within UNDP itself and with donors. Third, there is limited research about the conduct and quality of evaluations commissioned by multilateral organisations' country offices (Cida, 2006).

Evaluation methodology

A mixed method approach was adopted to assess the quality of UNDP's evaluations and the nature of its country level evaluation activities. An evaluation quality instrument was developed in order to assess the evaluations conducted at the country level. Two types of evaluation were assessed: outcome evaluations and project evaluations. Each evaluation was assessed against a set of evaluation criteria, each of which were considered to represent good evaluation practice. For each evaluation in the sample, each criterion was rated on a five point rating scale (Very poor, Poor, Fair, Good and Excellent).

Country offices in Uganda, Nepal, and Kenya were visited in order to appraise the evaluation activities at those offices. A broad range of evaluation stakeholders were interviewed. These included senior management at UNDP, programme officers and evaluation focal points at UNDP, evaluation consultants, partner government representatives, implementing agencies, and NGOs. Combined, their experiences of UNDP's evaluation activities provide a broad picture of how evaluation is being undertaken, its perceived value and issues that need to be further strengthened.

Key findings and conclusions

i) Quality of evaluations commissioned by UNDP's country offices

This evaluation found that, overall, evaluation quality was poor. Information was either missing, or the criteria were not applicable, in 25 per cent of the ratings. A further 25 per cent of the criteria were rated either as "Poor" or "Very poor". In other words, on an averaged basis, only half of all evaluations commissioned by UNDP's country offices are likely to be reliable as a basis for decision-making. This evaluation rated 20 per cent of the criteria as "Fair", which means that problems were present but that the criteria were nonetheless fulfilled satisfactorily. 30 per cent were rated "Good" or better. While some aspects of evaluation quality were stronger than others, the overall conclusion is that decision-makers should exercise caution when basing decisions about future programme activities on evaluations commissioned by the country offices. Overall evaluation quality does not meet a sound evidence basis for decision-making.

ii) Conduct and use of evaluations

Country office evaluation capacities

There is a general need for strengthened in-house evaluation capacities, in terms of both human and financial resources. In the case of Nepal, significant efforts have been undertaken in this respect. A proper monitoring and evaluation (M&E) unit has been established with two fulltime staff dedicated to M&E. This effort has a great impact on the position of M&E within the CO, and is a first step to ensuring good quality evaluation products and a strengthened evaluation culture. It is also a prerequisite for the CO engaging in policy dialogue with partner governments on issues related to M&E. Since there is not a sufficient accountability mechanism in place, senior management's interest will be crucial for delivering a strengthened evaluation focus at the COs.

Outcome evaluation – concept and methodology

The shift within UNDP to measuring outcomes of implemented projects and programmes has faced some challenges that require particular attention. The rapidly shifting and sometimes overlapping strategic policy documents appear to pose serious threats to the planning of evaluations and to conducting appropriate outcome assessments. This has led to COs manipulating their programmes to fit into over-arching predetermined outcomes. The current system favours a "backwards" way of working, in which outputs are subsequently linked to outcomes. In general, outcomes need to be clearly specified in order to be closely interlinked with project activities. Baselines are often missing, which further complicates proper assessments. The fact that outcomes (in most cases) can only be assessed several years after the termination of a programme indicates a change is required in programming to accommodate this reality. To maximise the results of an outcome evaluation, the outcome evaluation should be planned to be conducted several years after termination of the programme.

Evaluation partnerships

Advocating for the role and use of evaluations is core in order to achieve any improvement in national ownership of evaluation activities. Stakeholders' involvement in the evaluation process is important in order to improve the quality and utility of evaluations. Including national government representatives in the evaluation process has strong potential for ensuring a wider ownership of evaluations. Efforts should be made to providing government counterparts with better opportunities to effectively influence processes.

Use of evaluations

UNDP's track record for the effective and efficient use of (outcome) evaluation findings and recommendations as a tool for learning and programme improvements is not strong. This may be partly explained by the wide range of different strategic – often overlapping – documents that are relevant to individual country offices. Today, it is not clear the extent to which, and how, outcome evaluations feed into further activities. Hence, the very reason for undertaking these evaluations – assessing performance in terms of development results – appears to be being neglected.

The recently developed management response system within UNDP has potential for improving this situation, but it needs to be firmly anchored within the organisation. Hence, the adoption of clear incentives appear necessary in order to facilitate a constructive use of the system by COs. Also, evaluation reports need to be better disseminated and more clearly communicated in order to improve UNDP partners' use of evaluations.

iii) The strengthening of national evaluation capacities

UNDP’s mandate as a capacity-building agency in terms of M&E needs to be further explored. Findings from this evaluation indicate that M&E activities are not aligned with those of the partner country’s, and hence are not supporting the long-term development of M&E capacities. There should be continuous support to national M&E systems through the provision of financial and technical assistance. In cases where M&E tools and methodologies are developed, and capable staff exist within government, the skills of these people should be utilised.

Key recommendations

Below follows a summary of the main recommendations from this evaluation. These are further elaborated in Chapter 7.

1 Recommendation on general, and UNDP-specific, measures for improving the Swedish assessment of multilateral organisations’ performance

Since evaluation is an important element for providing improved performance information, Sweden’s approach for assessing multilateral organisations’ performance in terms of relevance and effectiveness could be further developed by including the following:

- An “in-house evaluation capacities ” check
- A “Country Office senior management accountability” check
- A “national capacity development and advocating for evaluation” check
- A “partnerships in evaluation” check
- A “reality check” of evaluation quality

2 Recommendations for a more coherent and strategic Swedish positioning on UNDP’s Executive Board

Principally, Sweden should:

- Advocate a simplified results reporting system

Furthermore, and to build on the results coming out of the assessments made by the Swedish Government of UNDP, it appears appropriate that Sweden advocates for improvements on the following issues:

- Advocate strengthened in-house evaluation capacities
- Advocate holding country office senior management accountable
- Propose M&E capacity building as a UNDP programmatic activity
- Propose advocating evaluation as a UNDP programmatic activity
- Advocate a strengthened role for the Executive Board in ensuring a correct implementation of the UNDP Evaluation Policy

3 Recommendations on issues to be addressed in Sweden's organisation-specific strategy for UNDP

It appears appropriate that Sweden's organisation-specific strategy for UNDP places a clear focus on strengthening the role and use of evaluation within the organisation. Hence, it is suggested that the issues addressed above are included in this strategy.

Furthermore, Sweden should consider allocating part of its non-core funding to activities that support the implementation of the Evaluation Policy at the country level. To the extent that such funding is provided to UNDP, it is proposed that some of these allocations focus on:

- The strengthening of in-house evaluation capacities
- Further development of the management response system
- The strengthening of national evaluation capacity building

Abbreviations

AfDB	African Development Bank
AWP	Annual Work Plan
CCA	Common Country Assessment
CCF	Country Cooperation Framework
CO	Country Office
CP	Country Plan
CPAP	Country Programme Action Plan
CPD	Country Programme Document
DG	Democratic Governance
EO	Evaluation Office (UNDP)
ERC	Evaluation Resource Center (UNDP)
INGO	International Non-Governmental Organisation
M&E	Monitoring and evaluation
MDG	Millennium Development Goals
MRE	Management Response to Evaluations
MYFF	Multi-Year Funding Framework
NGO	Non-Governmental Organisation
NIMES	National Integrated Monitoring and Evaluation Strategy
OE	Outcome Evaluations
OECD/DAC	Organisation for Economic Co-operation and Development/ Development Assistance Committee
PE	Project Evaluations
PEAP	Poverty Eradication Action Plan (Uganda)
PES	Programme Evaluation Standards
PRSP	Poverty Reduction Strategy Plan
RBA	Regional Bureau for Africa
RBAS	Regional Bureau for Arab States

RBAP	Regional Bureau for Asia & the Pacific
RBEC	Regional Bureau for Europe and the Commonwealth of Independent States
RBLAC	Regional Bureau for Latin America and the Caribbean
RBM	Results Based Management
UNDAF	United Nations Development Assistance Framework
UNDP	United Nations Development Programme
UNEG	United Nations Evaluation Group

Acknowledgements

We would like to thank the staff at the UNDP Evaluation Office in New York for supporting the conduct of this evaluation, and also the staff at the UNDP country offices in Kenya, Nepal and Uganda for their hospitality and the professional manner in which they organised the field visits. Finally, we would like to express our gratitude to all staff that were interviewed at the various organisations in Kenya, Nepal and Uganda for taking the time to meet with us and for providing us with the information we sought.

Contents

1 Rationale, objectives and purpose of the evaluation

1.1 Background – rationale of the evaluation

Through international commitments such as the Paris Declaration, international development cooperation is becoming more results-orientated. Evaluations have a crucial role to play in this process. At the same time, the Paris principles pose challenges to evaluation: evaluation should be nationally owned (ownership) and conducted in collaboration with other donors (harmonisation).

The focus on results in international development cooperation has had important implications for multilateral organisations. There is increasing pressure on these organisations from bilateral donors to improve their performance, and to make this clearly visible through monitoring, evaluation and reporting procedures.

At the same time, bilateral donors have developed their own tools for measuring multilateral performance. These approaches, using different methodologies and concepts of effectiveness, each have their strengths and weaknesses.[1]

In Sweden, the Strategy for Multilateral Development Cooperation (the Strategy) was approved by the Swedish Government in 2007. It has been developed as a first step towards clearer and more results-orientated Swedish work and involvement in multilateral development cooperation.[2] The Strategy proposes an assessment framework focusing on relevance and effectiveness. This assessment framework should be used as a guidance tool for financing decisions. It should also serve as a basis for developing specific strategies for the various organisations.

The emphasis on relevance and effectiveness in multilateral cooperation places heavy demands on follow-up and evaluation, and increases the expectations that results will be reported.[3] The Strategy stipulates that "the work of assessing an organisation's effectiveness must concentrate on results at the country level and consist chiefly of the multilateral organisation's own reporting."[4]

However, it should be recognised that although substantial efforts have been made in recent years, evaluations produced by multilateral institutions are still of very variable quality. As stated in the Strategy: "In the UN system, much remains to be done. Sweden should be active in making demands, push for improvements and ensure that experience gained is fed back into activities. Swedish resources should be made available as needed."[5]

[1] For a further description of different approaches for assessing multilateral effectiveness, see Appendix I.
[2] MFA of Sweden (2007): *Sweden's Strategy for Multilateral Development Cooperation*, Ch. 1.2, p 2.
[3] Idem, Ch. 3, p 14.
[4] MFA of Sweden (2007): Supra note 2, Ch. 2.2, p. 6.
[5] MFA of Sweden (2007), Supra note 2, p 15.

The general poor reporting on country level results from multilateral organisations is widely seen as a serious deficiency for appropriate assessments of multilateral performance.[6] Indeed, this should be seen as a serious problem, since development results are best assessed at the country level.

1.2 Objectives of the evaluation

This evaluation focuses on the United Nations Development Programme (UNDP). In 2006 Sweden channelled USD 107.1 million as core funding and more than USD 180.1 million as project support to UNDP. Sweden contributes more funding to UNDP than to any other UN agency. In providing this level of funding, Sweden is also one of the largest donors to UNDP.[7]

In order to achieve more results-orientated Swedish work with and involvement in multilateral organisations, the focus on follow-up and evaluation in the organisations that Sweden cooperates with must be strengthened. This evaluation's explicit focus is on evaluations commissioned by the country offices of UNDP. The first two objectives of this evaluation are to contribute to an improved knowledge of:

1 the quality of evaluations commissioned by UNDP's country offices, (are evaluations reliable as evidence of performance?); and

2 the conduct and use of these evaluations, (are evaluations conducted in a manner that ensures their reliability? Are they conducted in collaboration with other actors? Are evaluation results used in further programming?).

Related to these two objectives, the Swedish Strategy stipulates that: "It is also important to stimulate and develop the capacity of partner countries to perform domestic audit and evaluations of development initiatives. Such evaluations should be part of national follow-up and planning processes."[8]

Capacity building is an important measure for promoting a strengthened national ownership of evaluation. In order to achieve this effectively and efficiently, Sweden must increase its awareness of multilateral organisations' current efforts to strengthen national evaluation capacities. Hence, the third objective of this evaluation is to improve Sweden's knowledge about:

3 the efforts undertaken by UNDP in order to strengthen national evaluation capacities.

In terms of evaluation at the country level, UNDP is mandated to work with and through partner governments' M&E systems and support capacity development within partner countries. This is a feature of the organisation's mandate that sets it apart from other funds and programmes of the UN system, and therefore makes it particularly relevant as the subject of this evaluation.

[6] MFA of Sweden (2007), Supra note 2, p 14.
[7] In 2006, Sweden allocated USD 1148.3 million in multilateral core funding and USD 585.2 million as earmarked funding to multilateral organisations. *(MFA of Sweden (2007): UDs faktablad – Biståndet i siffror, 2007).*
[8] MFA of Sweden (2007), Supra note 2, p 14

Since this evaluation explicitly focuses on evaluations commissioned by UNDP country offices (hence excluding other evaluations and reports produced by UNDP), it is neither possible nor appropriate to formulate conclusions about whether UNDP's reporting is sufficient to enable a determination of the organisation's overall relevance and effectiveness, as required under the Swedish Strategy. It is nonetheless possible to highlight some important issues requiring particular attention for an improvement in the capacity for Sweden's reliance on evaluations produced by UNDP.

1.3 Purpose of the evaluation

Increasing the knowledge about evaluation quality and how UNDP is pursuing evaluations commissioned by its country offices, including the strengthening of national evaluation capacities, should provide valuable insights about problems that need to addressed, and what can be considered successful, in the performance of UNDP's current modality of managing for development results. Based on this information, this report provides recommendations about measures that would make Sweden's engagement in multilateral development cooperation more coherent and results-oriented.

The purpose of this evaluation is to provide:

1 Recommendation on both general, and UNDP-specific, measures for improving Sweden's assessment of multilateral organisations' performance,

2 Recommendations for promoting a more coherent and strategic Swedish role within the Executive Board of UNDP,

3 Recommendations on issues that should be addressed in Sweden's organisation-specific strategy for UNDP.

2 Evaluation scope and methodology

The focus of this evaluation is UNDP's country level evaluation activities. This focus is restricted to evaluation activities and evaluations commissioned at UNDP country offices. Therefore, only project and outcome evaluations undertaken by country offices are included, while evaluations and related activities of the UNDP Evaluation Office, regional bureaux and practice and policy bureaux fall outside the scope of this study.

This evaluation was initiated in February 2007 and finalised in June 2008. It covers UNDP country level democratic governance evaluations undertaken during the period 2004-2006, and takes relevant strategic policies related to evaluation activities into account.

The evaluation quality assessment under this evaluation was conducted without geographic delimitation. The random selection of evaluations commissioned by UNDP's country offices was drawn from UNDP's online evaluation resource centre (ERC), which carries a database containing evaluations uploaded by all UNDP country offices. Country office field visits were undertaken in December 2007 and January 2008. The assessment of evaluation systems in this evaluation is limited to case studies of the country offices in Kenya, Uganda and Nepal.

A mixed method approach, using an "evaluation quality assessment" and an interview-based "evaluation system assessment", was adopted to assess the overall evaluation quality in UNDP and the nature of its country level evaluation activities. Triangulation between the two approaches was carried out, mainly to compare and cross-reference findings. This was also useful for identifying factors that both inhibited and facilitated evaluation activities.

2.1 Evaluation quality assessment

This evaluation only considers UNDP evaluations conducted in the area of democratic governance, which is the most important practice area of UNDP in terms of programme expenditures.[9] The evaluation was delimited to focusing on one practice area of UNDP in order to retain clear focus according to the study parameters. The sample of evaluations was randomly selected from UNDP's ERC database, and included evaluations undertaken in the period 2004-2006. From a population of 150 democratic governance evaluations, 25 evaluations (17 per cent) were selected for assessment. This constituted a sample from the ERC that was adequately representa-

[9] Almost half of UNDP total programming expenditures were dedicated to governance. In terms of funding sources, of total expenditures to the democratic governance practice area, 9 per cent came from regular resources, 45 per cent was "other resources from bilateral and multilateral donors", while the remainder, 46 per cent, constituted "other resources from program country governments" (UNDP (2007): *Multi-year funding framework cumulative report on UNDP performance and results for 2004-2006*, Annex. Programme expenditure, 2004-2006, table 1 and 2).

tive of outcome/project evaluations, service lines[10] and geographic coverage. The sample of evaluations is attached in Appendix II.

Each evaluation in the sample was assessed against 10 different aspects of evaluation quality.[11] Each of the 10 different aspects of quality was underpinned by a unique group of criteria considered relevant for that particular aspect of quality (see Appendix III, which describes the evaluation quality assessment instrument, for further details about how the criteria relate to the aspects of quality). For each evaluation, each criterion was rated against a 5-point scale. It should be noted that all criteria were attributed the same weight in the overall assessment. This could be considered an analytical weakness, as certain criteria are clearly more important for "good quality" than others (for example, "evaluation relevance" is clearly more critical than "report is free from grammatical errors"). However, since the analysis is structured around different aspects of quality it is possible to detect the loci of the most critical weaknesses in terms of quality. The assessed evaluations were registered in a matrix, permitting conclusions on different aspects of evaluation quality in the sample.

The analytical framework required the development and definition of the quality assessment instrument, and the calibration into "appropriate" values of each step in the scale for each indicator. The Programme Evaluation Standards[12] were used as a reference point to ensure consistency with international evaluation standards. Standards reflecting and assessing "the utility of evaluation" were given priority. In addition, the DAC Evaluation Quality Standards[13] and the UNEG norms and standards for evaluation in the UN system[14] were consulted.[15] The goal was to design an instrument that would be flexible enough to apply universally, rather than be limited to a specific organisation or context.

The method and process for defining the different values in the scale, and for calibrating the assessment, is described in the box below:

---

[10] The democratic governance practice area is divided into seven different "service lines": i) Decentralisation, local governance and urban/rural development, ii) E-governance and access to justice, iii) Electoral systems and processes, iv). Justice and human rights, v) Parliamentary development, vi) Policy support for Democratic Governance, and vii) Public administrator reform and anti-corruption.

[11] The 10 evaluation quality aspects applied in the instrument are: 1. Overall structure of the report, 2. Description of purpose, objectives and scope, 3. Description of evaluator credibility, evaluation independence and ethics, 4. Description of information sources, 5. Intervention description, 6. Evaluation report relevance, 7. Evaluation methodology, 8. Assessment of evaluation findings, 9. Use of OECD/DAC evaluation criteria, 10. Completeness of the report.

[12] Sanders, J. R.(1994): *The Programme Evaluation Standards*, 2nd edition, The Joint Committee on Standards for Educational Evaluation.

[13] OECD DAC (2002): *Evaluation Quality Standards*, OECD.

[14] UNEG (2005): *Norms and Standards for Evaluation in the UN System*, UNEG, New York.

[15] Forss et. al.(2008): *Are Sida evaluations good enough? An assessment of 34 evaluation reports*, A study commissioned by Sida, Stockholm. The methodology used in that evaluation was also considered in the development of the quality assessment tool here.

Box 1: Assuring consistency in the quality assessment: The evaluation team consisted of two evaluators. The set of criteria was tested on three randomly selected evaluations. This led to modifications and changes, involving adding and deleting. The evaluators then independently read the same three randomly selected evaluations. The ratings made by the two evaluators were compared in order to clarify discrepancies in rating and for establishing guidelines and common definitions based on the literature. A second random selection of another three evaluations was also read independently by the two evaluators, and thereafter compared and discussed in the same manner. The objective of this two-stage procedure was to ensure consistency of the data reported in this study. Also, through this exercise, a guide to the quality assessment instrument was developed, defining the requirements for the 1 – 5 rating of each criteria. After the evaluators had read, rated and normalised their respective ratings, the evaluators proceeded to the next phase of the evaluation analysis. The entire sample of evaluations in the study was divided equally between the two evaluators, for rating on the basis of the guide that was developed in the first phase of the analysis. Every fifth evaluation was read by both evaluators, independently, in order to ensure consistency and as a control mechanism for assuring that ratings were maintained in accordance with established definitions.

The quality assessment instrument and related guide can be found in Appendices III and IV.

2.2 Evaluation system assessment

The evaluation system assessment included interviews with a wide range of stake-holders of UNDP's country level evaluation activities. Interviews were conducted with the UNDP Evaluation Office, UNDP staff at three country offices and with evaluation stakeholders in these countries. Independent (international and national) consultants who had carried out evaluations commissioned by UNDP were also interviewed.

The main interview techniques were semi-structured interviews, both individual and in focus groups. The interview guides were organised around a number of issues as defined in the Terms of Reference for the country office field studies:

1. Aspects affecting evaluation quality
   - Evaluator competencies
   - Time spent on evaluation
   - Programme design and evaluation prerequisites
2. The conduct and use of evaluations
   - In-house evaluation capacity and institutional arrangements
   - Outcome evaluation: concepts and methodologies
   - Evaluation partnerships

   - Evaluation use

3. Strengthening of local evaluation capacities

   - Evaluation in national systems – measures for supporting national evaluation capacities

The Terms of Reference for the country office field visits, the interview schedules and the interview guide is included in Appendix V, VI and VII.

2.2.1 The selection of country offices for field visits

The selection of country offices for inclusion into the present evaluation was based upon: i) how far, as indicated by the UNDP Evaluation Office,[16] the country offices have progressed in the transition towards measuring outcomes instead of outputs, ii) the number of outcome and project evaluations submitted to the ERC,[17] and iii) the country context in which the country office is operating.

The country offices in Kenya, Uganda and Nepal were selected for inclusion in the present evaluation. In consideration of the criteria described above, the original proposal to retain a strict focus on Africa in the sample of country offices was aborted. This choice was based on the notable differences in evaluation activities between different regions.[18] It was therefore considered that such a narrow focus might have prevented the possibility of drawing generally relevant conclusions.

To deliver effectively on its mandate to work in direct partnership with governments, it is critical that UNDP adapt to the country context in which it operates. Therefore, it was considered enriching for the objectives of the present evaluation to select a number of countries with clear differences in development and political contexts.[19]

Although based on a small sample (three of a total of 133 country offices), observations made in the three country offices visited provide indications of a more general nature concerning constraints, challenges and enabling factors facing country offices in their evaluation activities. Obviously, a larger number of country offices included in the present evaluation would have enriched and reinforced the conclusions that are drawn from these visits. However, present resource and time constraints necessitated this smaller, but nonetheless quantitatively meaningful sample.

For a brief description of the country contexts in the countries visited, see Appendix VIII.

2.3 Information sources

A wide range of information sources (documents produced by UNDP, other multilateral organisations, a wide range of bilateral donors and academic researchers) have

[16] As indicated by the UNDP Evaluation Office when interviewed, October 2007.
[17] For further details, see Ch. 3.1.1.
[18] The average number of evaluations conducted by country offices in 2006 was 1.68 in the Asia Pacific region, and only 0.84 in the Africa region. (In the Asia Pacific region, with a total of 25 Country Offices, 42 evaluations were conducted, while in the Africa region, with a total of 45 country offices, 38 evaluations were carried out.) Seven per cent of evaluations in the Asia Pacific region were outcome evaluations, while in the Africa region 13 per cent were outcome evaluations. (UNDP (2007): *Annual report on evaluations in UNDP in 2006*).
[19] For further details, see Appendix VIII.

been used in the present evaluation, in particular for the background study. Information has been cross-validated and critically assessed through: i) triangulation of findings from the evaluation quality assessment with the evaluation system assessment, and ii) crosschecking of different stakeholder group interviews.

A complete list of information sources is provided at the end of this report. Core information used for this evaluation consists of:

- A sample of evaluations submitted to and held by the ERC;
- Literature and other documentation on meta-evaluation methodology;
- Literature and other documentation on bilateral donors' approaches for assessing the performance of multilateral organisations;
- Literature and other documentation concerning UNDP, its evaluation function and the process for becoming a results based management (RBM) organisation;
- Interviews with UNDP staff at UNDP headquarters (Evaluation Office) and at selected country offices, with evaluation stakeholders (partner countries' governments, implementing agencies, CSOs and Swedish MFA/Sida staff in the field) and with independent (international and national) consultants. A list of interviewees is included in Appendix IX.

2.4 Relevant stakeholders consulted

An important number of different stakeholder groups have been identified and, to the extent possible, involved in the evaluation process. These include, primarily, representatives from the Swedish Ministry for Foreign Affairs. This evaluation process also involved the UNDP Evaluation Office and the UNDP Nordic Office in Sweden.

2.5 Independence, ethics and quality assurance

Evaluation independence

SADEV is a government agency that independently initiates and conducts evaluations aiming to achieve a better, more efficient and effective Swedish development cooperation. Hence, the evaluators in this present study are independent from policy, operations and management functions of the Swedish MFA/Sida, UNDP and interviewees.

The evaluation team has been able to work freely and without interference. The cooperation of the UNDP Evaluation Office and participating country offices was greatly appreciated by the evaluation team, and that cooperation facilitated the work of the team. These entities displayed openness and on the whole granted permissions willingly to access relevant information.

Evaluation ethics

The evaluation has been undertaken by the evaluation team with integrity and honesty. Interviewees are not cited in this report and have been provided with the opportunity to review the findings drawn from the interviews.

Evaluation quality assurance

The main stakeholders have been given the opportunity to comment on findings, conclusions and recommendations. Factual errors have been amended, and other comments have been considered and taken into account to the extent that they were considered valid and relevant by the evaluation team. SADEV is fully responsible for the content of this final report.

The DAC Evaluation Quality Standards[20] have guided the structuring and drafting of this report.

As regards the quality assessment of a sample of evaluations, a continuous quality control has been exercised internally within the evaluation team as described in Box 1, Chapter 2.1.

Quality controls have been exercised during the evaluation process through internal seminars at SADEV. A reference group has provided comments and advice on draft evaluation plans and methodological issues.[21]

[20] OECD DAC (2002), Supra note 13.

[21] The authors would like to thank Jan Cedergren, Kim Forss, and Howard White for their thoughtful comments and suggestions throughout the evaluation process that ultimately advanced the quality and analytical depth of the report. These persons are not responsible for the information, content, conclusions and recommendations contained in this evaluation report.

3 UNDP – Strategic framework, the evaluation function and RBM

3.1 Core planning and strategic instruments

Evaluation within UNDP, at both the centralised and the decentralised level, is guided by a number of corporate strategies and planning instruments. These set the parameters for UNDP regarding how it should operate internally and in relation to partners, and influence the manner in which all evaluation activities are planned and carried out.

The Strategic Plan (2008-2011) is UNDP's corporate planning instrument, outlining the vision and mission of the organisation, as well as the concrete goals and objectives to be pursued over a three year cycle to support programme countries in achieving national development objectives. The purpose of the Strategic Plan is to improve the manner in which development outcomes and institutional results are achieved. It provides indicators to facilitate planning and monitoring.[22]

The common strategic framework for the operational activities of the United Nations system at the country level, the United Nations Development Assistance Framework (UNDAF) is intended to provide a collective, coherent and integrated United Nations system response to national priorities and the needs within the framework of the Millennium Development Goals (MDGs) and other international conventions and commitments. The UNDAF emerges from the analytical and collaborative efforts of the Common Country Assessment (CCA), and is the foundation for United Nations system programmes of cooperation.

The UNDP-specific Country Programme Document (CPD) and the Country Programme Action Plan (CPAP) are developed simultaneously with the UNDAF. CPDs are approved by the UNDP Executive Board and provide further definitions of the UNDAF outcomes as related to UNDP activities. CPDs are also required to take the cross-cutting issues, as defined in the UNDP Corporate Strategic Plan, into account. The CPAP, which is the document that makes the priorities of the CPD operational, sets the targets and goals for UNDP's cooperation with the partner country government and defines the parameters upon which the cooperation should be based. In terms of evaluation, the formulation of the CPD is accompanied by the development of an evaluation plan which is intended to set the stage for national ownership and alignment of UNDP evaluation with national systems.

3.1.1 Application of strategic documents at the country offices visited

As indicated in Table 1, below, UNDP's comprehensive corporate planning instruments (the Multi-Year Funding Framework (MYFF) (2004-2007) and, since 2008, the

[22] UNDP: *User Guide, Results Management and Accountability*: http://content.undp.org/go/userguide

Strategic Plan (2008-2011)) apply to all country offices. The UNDAFs of the three countries under this study do, however, cover different periods of time: in Kenya, the UNDAF 2009-2013 (with related CPAP) enters into force one year after the Strategic Plan. In Uganda the current UNDAF (with related CPAP ending one year before the UNDAF) covers the 2006-2011 period. In Nepal, the UNDAF (with its related CPAP) enters into force at the same time as the Strategic Plan. The evaluation plans established by the country offices in Uganda and Nepal cover the same period of time as the CPAP, while in Kenya the present evaluation plan covers 2006-2008 (the current CPAP covering 2004-2008).

Since the tools defining outcomes for a country programme (MYFF/Strategic Plan and the UNDAF) sometimes overlap, assessing progress towards outcomes appears not to be a straightforward task. There is hence no single, consolidated, clearly articulated and logical framework for outcomes and contributing outputs that governs how the UNDP country offices should operate.

Kenya has submitted three outcome and three project evaluations to the ERC, Uganda two outcome and four project evaluations, and Nepal four outcome and 17 project evaluations.

Table 1 Applying strategic document, evaluation plan and the number of outcome/project evaluations submitted to the ERC.

	MYFF	Strategic Plan	UNDAF (1)	CPAP (1)	UNDAF (2)	CPAP (2)	Eval. plan	OE in ERC	PE in ERC
Kenya	2004 - 2007	2008 - 2011	2004 - 2008	2004 - 2008	2009 – 2013	2009 - 2013	2006 - 2008	3	3
Uganda	2004 - 2007	2008 - 2011	2001 - 2005	2001 - 2005	2006 – 2011	2006 - 2010	2006 - 2010	2	4
Nepal	2004 - 2007	2008 - 2011	2002 - 2006	CCF 2002 – 2007*	2008 – 2010	2008 - 2010	2008 - 2010	4	17

* Country Cooperation Framework - Extended for another year to 2007 due to a programme review conducted by UNDP and the GoN. The extension was primarily done to reprioritise and refocus projects as a consequence of the resolution of the conflict in Nepal.[23]

3.2 The evaluation function in UNDP

UNDP has a long-standing tradition of emphasising the importance of monitoring and evaluation. Evaluation has long been perceived as an important issue and the organisation has recognised that improvements are necessary in order to strengthen the role of evaluations in programming. The Executive Board approved the latest Evaluation Policy in 2006 in the context of this ongoing support.

The evaluation function within UNDP is characterised by a centralised component: the Evaluation Office (EO) reporting to the Executive Board through the Administrator; and a decentralised component, evaluation activities undertaken by the country offices and other organisational entities.

[23] UNDP Nepal (2008): *Country Programme Action Plan* 2008-2010.

The 2006 Evaluation Policy provides the basis for all evaluation activities within UNDP and its associated funds and programmes. The policy stipulates the key principles underpinning all evaluations, as:

> National ownership, human development and human rights, coordination and alignment in the United Nations system, and managing for results.[24]

The purpose of evaluation in UNDP is mainly to "… *address what works and why, as well as what does not work and unintended outcomes. This will support accountability, inform decision-making and allow UNDP to better manage for development results.*"[25] As it is in most organisations that use evaluation, evaluation in UNDP is concerned with supporting accountability and learning about how to improve future programming.

3.2.1 Centralised level – the Evaluation Office (EO)

Evaluations undertaken by the EO are intended to serve the purpose of enhancing corporate accountability, strategic planning, and the development of information for global knowledge use. The EO is mandated to: i) undertake strategic evaluations of UNDP management and programme policies, ii) promote the use of evaluation findings and support knowledge management, iii) develop evaluation guidelines and methods, and iv) provide general support for an improved results orientation in the organisation.[26]

The EO has obligations towards the decentralised component of the evaluation system. These include: i) maintaining a dialogue with programme countries and country offices to strengthen the internalisation and utilisation of evaluation standards, both within UNDP and amongst its partners, and ii) strengthen programme country evaluation capacity and involvement in evaluations through country-led evaluations, joint evaluations and use of partner country professional resources.[27]

Box 2: Independent evaluations conducted by the Evaluation Office include:[28]

- *Strategic evaluations* that assess UNDP performance in areas that are critical to ensuring sustained contribution to development results in the context of emerging development issues and changing priorities at the global and regional level.
- *Programmatic evaluations*
  - *Global, regional and South-South programme evaluations* that assess the performance and intended and achieved results of those programmes.
  - *Assessments of Development Results (ADR) that assesses the attainment of intended and achieved results as well as UNDP contributions to development results at the country level.*

[24] UNDP (2006): *The evaluation policy of UNDP*.
[25] Idem.
[26] UNDP: Supra note 24
[27] UNDP: Supra note 24
[28] UNDP: Supra note 24

3.2.2 Decentralised level – country office evaluation activities

UNDP is a highly decentralised organisation, present in 166 countries with a total of 133 country offices. Country offices, regional bureaux, and practice and policy bureaux commission evaluations within the programmatic frameworks for which they are responsible. The focus of decentralised evaluations is on information for programme improvement and the development of new programmatic frameworks. Moreover, evaluations commissioned by UNDP's country offices should provide the EO with information for higher level evaluations. The structure of the organisation does not, however, prevent the decentralised part of the evaluation system from being largely managed in a centralised manner. Important decisions are being made at UNDP headquarters that will affect how evaluations are conducted at the country level. For instance, country programme documents are approved by the Executive Board, the decision to use outcome evaluations was made at headquarters, and the corporate model for RBM has obvious effects on how country offices are managed.

Senior management at country offices are responsible for ensuring that evaluation activities are pursued in a professional manner and consistent with the UNDP Evaluation Policy. In this context, senior management should ensure that evaluation findings are being utilised in order to improve the quality of programmes, to guide strategic decision-making on future programming and positioning, and to share knowledge on development experience. As stipulated by the Evaluation Policy, the main responsibilities of senior management include: i) identifying, with partner governments and key stakeholders, priority areas for evaluation when preparing programmes, and in designing and implementing a strategic evaluation plan, ii) establishing an appropriate institutional arrangement to manage evaluation, iii) ensuring adequate resources for evaluation, and iv) preparing management responses to all evaluations.[29]

Box 3: The main decentralised evaluations are:

- *Outcome evaluations* that address the short-term, medium-term and long-term results of a programme or cluster of related UNDP projects. They include an assessment of the effectiveness, efficiency, sustainability and relevance of the programme against their objectives, their combined contribution, and the contribution of external factors and actors.

- *Project evaluations* that assess the effectiveness and the efficiency of a project in achieving the intended results. They also assess the sustainability and relevance of outputs as contribution to medium-term and longer-term outcomes.

- Additionally, UNDP Country Offices may decide to commission other types of evaluations such as UNDAF joint evaluations, in collaboration with other UN agencies, and the evaluation of their country programme.

Outcome evaluations identified in the evaluation plan are mandatory, while project evaluations are mandatory when required by a partnership protocol.

[29] UNDP: Supra note 24

Monitoring and evaluation of the CPAP is undertaken in accordance with the UNDAF results matrix and monitoring and evaluation plan. The planning for monitoring and evaluation at the country level is intended to be an integral part of the programme planning process. Country offices make the evaluation plan available as an annex to all CPDs that are submitted for approval to the Executive Board. During the CPAP preparation, country offices elaborate the plan further, in consultation with national stakeholders. Selected country programme outcomes will be evaluated once during the cycle. All country programmes will be subject to an evaluation within the final year of their cycle.

3.3 Evaluation and Results-Based Management (RBM)

Many organisations within the United Nations system, including UNDP, have embarked on a process of introducing performance management systems, referred to as Results-Based Management (RBM). RBM has been defined as a broad management strategy aimed at achieving improved performance and demonstrable results. A recent evaluation on RBM within UNDP provided a set of conclusions that were generally critical of aspects of the implementation of RBM in the organisation.[30] The evaluation's conclusions included: i) there is a weak culture of results in UNDP, ii) managing for results requires strengthened leadership at several levels in the organisation, and iii) RBM has been misinterpreted as not supporting the decentralised way in which UNDP works.

The role of evaluation in relation to RBM can be viewed in different ways. A recent survey undertaken by UNEG captures the perspectives of the evaluation offices in a number of UN agencies.[31] The survey concluded that there is:

> … a general agreement among respondents that evaluation has received a new lease of life with the introduction of RBM, and that, in particular, RBM has strengthened the role of evaluation, facilitated the promotion of an evaluation culture, enhanced the use of evaluation findings in programming and put greater emphasis on self-evaluations by program managers.[32]

The survey further evidenced mixed opinions about whether evaluation should be considered part of RBM. Two main opinions were evident in the respondents of the survey: some regarded evaluation as an integral component within the RBM conceptual framework, while others considered evaluation to be a more subsidiary, but nonetheless complementary function. Nevertheless, given its focus on results and methodologies for assessing performance, evaluation is, undoubtedly interlinked with the concept of RBM.

Results Based Management at UNDP has been highly associated with the concept of the Multi Year Funding Framework (MYFF), which was first introduced to the organisation in 2000. The MYFF was the key corporate planning instrument, which

[30] UNDP (2007): Evaluation of Results-Based Management at UNDP.
[31] UNEG (2007): *The Role of Evaluation in Results-Based Management (RBM).* The UNEG survey on the role of evaluation in RBM in UN organisations was sent to 26 UN entities, including specialised agencies, the UN Secretariat and funds and programmes. By 31 March 31 2006, 21 of 26 UN organisations surveyed had responded to the questionnaire.
[32] Idem.

was subsequently replaced by the Corporate Strategic Plan in 2007. Between 2000 and 2007 there were two rounds of the MYFF, before the instrument was abandoned. At the core of the MYFF was the idea of integrating programme objectives, resources and outcomes within corporate priorities and focus. The introduction of Results Based Management (RBM) led to a shift of focus, from inputs to outputs and outcomes, which ultimately promoted further alignment between country programmes and UNDP corporate strategic goals.

In the context of RBM, outcome and project evaluations should generate information about UNDP's contribution to development results for the purposes of learning, line oversight and public accountability at country, regional and global levels.[33]

The scope of outcome evaluations is generally broad, taking into consideration a wide range of actors, projects and programmes when assessing what contributed to a specific outcome. The purpose of outcome evaluations is to enhance development effectiveness by assisting decision-makers and by redirecting interventions.

[33] UNDP: Supra note 24.

4 Main findings from the quality assessment of evaluations

This section discusses the main findings from the quality assessment of the sample of evaluations commissioned by UNDP's country offices. The analysis is structured around 10 sub-headings that individually capture different aspects of evaluation quality. The number of criteria under each sub-heading ranges from 3 to 7. Significant strengths and weaknesses in terms of evaluation quality are detected and further commented on in the analysis.

4.1 General information about the evaluation sample

Table 2 General information about the evaluations assessed.

	Project	Outcome								
Evaluation Type	16	9								
	2004	2005	2006							
Year	9	7	9							
	RBEC	RBLAC	RBAP	RBA	RBAS					
Region	8	3	4	9	1					
	1 week	2 weeks	3 weeks	4 weeks	5 weeks	>5 weeks	$ info	No info		
Cost of/time spent on evaluations	0	7	4	4	0	3	1	6		
	UNDP alone	UNDP +gov	1 co-fin	2 co-fin	3 co-fin	4 co-fin	5 co-fin	>5 co-fin	UNDP not financing	Unclear
Financing of intervention	5	4	2	2	1	1	0	5	3	2
	Survey	Interviews	Observations	Document Analysis	Other					
Methods Used	3	23	14	24	1					

Of the sample of 25 evaluations, 16 were project evaluations and nine were outcome evaluations. The evaluations were conducted in the 2004-2006 period and covered all geographic regions in which UNDP is present.

The relative scarcity of information regarding the actual costs associated with undertaking individual evaluations was a principal factor that was evident in reviewing the evaluation sample. Rarely was any information provided in this regard. UNDP undoubtedly keeps records of costs, but these statistics are not transparent for an outside reviewer. This information is essential in order to better understand the importance of evaluation in relation to overall costs for the evaluated activity. Nonetheless, in the absence of this information, an able substitute indicator is the amount of time dedicated to undertaking an evaluation. The average time spent on undertaking the evaluations was 3.25 weeks.

In reviewing the methods and data collection instruments that were applied in the evaluations, it was evident that most evaluations followed a standard procedure. The main instruments included interviews, observations (such as site visits) and document analysis. In 14 of 25 reports, the evaluators undertook more thorough observations than just interviews and document analysis. It is important to highlight the absence of surveys or other types of data collection tools that could provide a quantitative foundation for analysis in the evaluations analysed. This absence is perhaps attributable to the fast pace at which most evaluations are undertaken, as mentioned above. More in-depth analysis and complex methodologies require a longer time-frame.

4.2 Overall quality of assessed evaluations

Figure 1 Summary of ratings for 25 evaluations on all criteria

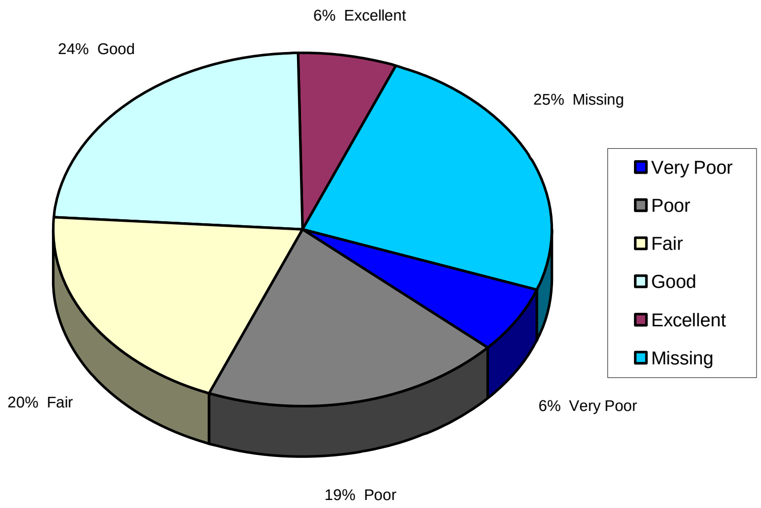

Figure 1, above, displays a summary of all 25 evaluation reports, assessed against the 49 criteria which each represent elements of good practice. Information was either missing (that is, it was impossible to rate according to the criteria) or was not applicable (that is, the content of the criteria did not relate to the evaluation) for 25 per cent of the criteria (that is, it was not possible to rate to 25 per cent of the criteria overall). A further 25 per cent of the criteria were either rated "poor" or "very poor" (19 per cent; six per cent "very poor"), which means that, on an averaged basis, only half of all evaluations undertaken at the country level were of reliable quality in serving as a basis for decision-making. Twenty per cent of the criteria were rated "fair", which means that certain problems were present but that the criteria were fulfilled satisfactorily. 30 per cent of the criteria were rated "good" or better.

4.3 Assessed aspects of evaluation quality

Figure 2 Summary of ratings for 25 evaluations on all criteria grouped to display different aspects of evaluation quality

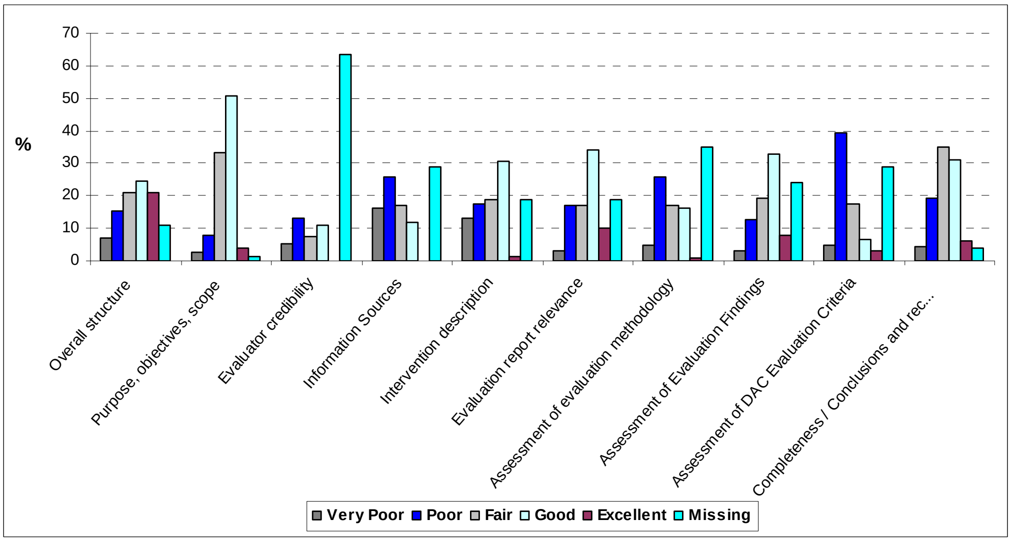

Figure 2, above, shows the assessments in relation to different aspects of evaluation quality, identifying areas in which improvements are necessary in order to improve overall quality.

4.3.1 Overall structure of the reports (7 criteria)

		Very Poor	Poor	Fair	Good	Excellent	Missing
	Overall structure of the report (in percentages)						
1	The report is brief, simple and direct	0	20	24	44	12	0
2	There is a clear and logical structure to the chapters of the report and to the report as a whole	4	16	24	40	12	4
3	There is a clear and adequate executive summary	0	8	24	20	12	36
4	The annexes are well structured and readable	0	8	20	32	16	24
5	The report makes use of references and these are appropriately documented	36	24	24	12	0	4
6	The report is free from grammatical and spelling errors	0	0	4	4	92	0
7	Data collection instruments are carefully described in the evaluation or its annexes	8	32	28	20	4	8
	Summary of ratings	**7**	**15**	**21**	**25**	**21**	**11**

The criteria for "overall structure of the reports" should determine how well the evaluations are structured in terms of clarity and directness. In order for an evaluation to be used, it must be understood. The 25 evaluations received relatively high scores on these criteria. Forty six per cent of the criteria in this section were rated either "good" or "excellent" (25 per cent; 21 percent respectively). 21 per cent were rated "fair" and 22 per cent were rated either "poor" or "very poor". The majority of evaluations assessed were relatively well structured and clear.

However 36 per cent of the evaluations (9 reports) were missing executive summaries. Moreover, 24 per cent (6 reports) did not have any annexes attached, notwithstanding that reference was made to these within the reports. In these cases further assessments of the evaluation reports became problematic, as assessing and validating much of the analysis hinged on data provided in the annexes. Most evaluations scored poorly in using references satisfactorily. References to scientific journals and other research are largely absent. The criterion which controlled for grammatical accuracy scored 92 per cent. The assessment of this criterion considerably boosted the overall rating for "overall structure".

4.3.2 Definition of purpose, objectives and scope of the evaluation (3 criteria)

		Very Poor	Poor	Fair	Good	Excellent	Missing
	Purpose, objectives, scope (in percentages)						
8	The purpose of the evaluation are described in enough detail, so that it can be assessed	4	4	32	56	4	0
9	The objectives of the evaluation are specified	4	4	32	52	4	4
10	The scope of the evaluation is clearly defined	0	16	36	44	4	0
	Summary of ratings	**3**	**8**	**33**	**51**	**4**	**1**

It was not uncommon that the reasons or rationale for undertaking an evaluation was absent from the evaluation report, or that these were described vaguely and lacked focus. This set of criteria is aimed at determining the extent to which the evaluation is clear in this regard and whether the evaluation findings are intended to be used in a subsequent policy process or project design phase. The sample of evaluations scored highly in this aspect. In most cases the purpose and objectives were obvious, as the evaluation would determine whether a project should be allowed to continue or not. Fifty four per cent of the criteria in this section were rated either "good" or "excellent" (51 per cent; four per cent respectively). Thirty three per cent were rated "fair" and 11 per cent were rated either "poor" or "very poor". As Figure 2 illustrates, this part of the assessment received the highest ratings.

The assessments of individual criteria shows that evaluation scope received lower ratings than purpose and objectives.

4.3.3 Evaluator credibility, evaluation independence and ethics (7 criteria)

		Very Poor	Poor	Fair	Good	Excellent	Missing
	Evaluator credibility, evaluation independence and ethics (in percentages)						
11	There is an indication of the level of impartiality in reporting	8	8	0	0	0	84
12	The evaluation team possesses a mix of evaluative skills and thematic knowledge	16	28	12	24	0	20
13	The report discusses evaluation ethics	0	0	0	4	0	96
14	There is an indication of the level of independence in reporting	4	20	0	4	0	72
15	There is a discussion on threats to reliability and validity	0	12	4	4	0	80
16	Conflicts of interest are addressed openly and honestly. Measures have been taken to avoid conflict of interests	4	4	0	4	0	88
17	Stakeholders, that is, persons involved in or affected by the evaluation are identified and have been involved throughout the process	4	20	36	36	0	4
	Summary of ratings	**5**	**13**	**7**	**11**	**0**	**63**

Information about the composition of evaluation teams and the level of independence that each evaluation team enjoyed is crucial, as these are fundamental aspects of good evaluation quality. In assessing the 25 evaluation reports against the seven criteria covering this aspect of quality, information was missing or not taken into consideration in 63 per cent of the criteria overall. While these criteria could be regarded as difficult to “describe and develop” in the evaluation report, nonetheless it could be argued that such information should be provided in the terms of reference for the evaluation. In any event, the fact that such a large percentage of data is missing is an area that should be addressed by UNDP.

The criteria checking for information on ethics and potential conflict of interest were rated missing in 96 per cent and 88 per cent of the sample, respectively. The crucial criterion which controlled for the extent to which stakeholders have been participating in and contributing to the evaluation process received more positive scores. 36 per cent (9 reports) were rated either “fair” or “good” in this regard. This criterion takes into account whether the evaluations provided specific information about stakeholder involvement and in what parts of the evaluation process they made an active contribution. The assessment did not go into further detail about means and modalities of the elements of participation and contribution.

4.3.4 Information sources (4 criteria)

		Very Poor	Poor	Fair	Good	Excellent	Missing
	Information Sources (in percentages)						
18	The sources of information used are described in enough detail	12	48	24	16	0	0
19	All potential information sources appear to be exhausted	12	40	20	24	0	4
20	The evaluation cross-validates and critically assesses the information sources used	32	12	16	4	0	36
21	There is a discussion about the limitations of information	8	4	8	4	0	76
	Summary of ratings	**16**	**26**	**17**	**12**	**0**	**29**

The manner in which information sources are displayed and reported is crucial in determining the accuracy and utility of an evaluation. It is important that the evaluation remains critical towards the information sources used so that accurate inferences can be made from the data collection process. In assessing the 25 evaluation reports against the four criteria covering this aspect of quality, the overall rating was quite low. No information at all was provided for 29 per cent of the criteria overall in this section. It should be possible to apply all of the criteria, as there is no legitimate reason why this information could not be included in the reports . Moreover, 42 per cent of the criteria were rated as either “very poor” or “poor” (16 per cent; 26 per cent respectively).

Overall, the sources of information were described unsatisfactorily. 60 per cent of the evaluations (15 reports) were rated “very poor” or “poor” in this aspect (12 per cent; 48 per cent respectively). Also, few evaluations attempted to discuss the limitations of the information sources used and what this implied for the validity of evaluation findings. 76 per cent of the evaluations (19 reports) contained no such discussion, nor any reference to the fact that information sources might have limitations.

4.3.5 Intervention description (3 criteria)

		Very Poor	Poor	Fair	Good	Excellent	Missing
	Intervention description (in percentages)	0	0	0	0	0	0
22	The intervention and its context are described	0	12	32	52	4	0
23	The intervention’s side effects are identified	20	20	8	8	0	44
24	The organisational arrangements established for implementation of the intervention is described	20	20	16	32	0	12
	Summary of ratings	**13**	**17**	**19**	**31**	**1**	**19**

Evaluations should include a description of the intervention being evaluated. In order to assess the accuracy of the evaluation analysis it is important that there is an understanding of what is being evaluated, and its unique features and components. This quality aspect includes three criteria that assess how the intervention has been described. The interventions were generally well described in the sample reports. 50 per cent of the criteria overall in this section were rated either “fair” or “good” in this aspect (19 per cent; 31 per cent respectively).

4.3.6 Evaluation report relevance (4 criteria)

		Very Poor	Poor	Fair	Good	Excellent	Missing
	Evaluation report relevance (in percentages)						
25	The terms of reference are clear and focused	0	12	12	32	12	32
26	The report is focused on the tasks as defined in the terms of reference	4	12	20	32	4	28
27	The evaluation question(s) is/are clearly stated	4	20	28	24	16	8
28	The report is focused on the evaluation questions	4	24	8	48	8	8
	Summary of ratings	**3**	**17**	**17**	**34**	**10**	**19**

The four criteria that aim to capture evaluation report relevance focus broadly on the clarity of evaluation questions and the extent to which the evaluation corresponds with its terms of reference. Overall, 44 per cent of the criteria were rated either “good” or “excellent” (34 per cent; 10 per cent respectively).

32 per cent of the sample (eight reports) were rated “good” in terms of clear and focused Terms of Reference, while 32 per cent of the sample (eight reports) had no Terms of Reference attached. The criterion that treated focus on evaluation questions was rated “good” in 48 per cent of the sample (12 reports) and “excellent” in eight per cent (two reports).

4.3.7 Evaluation methodology (4 criteria)

		Very Poor	Poor	Fair	Good	Excellent	Missing
	Assessment of evaluation methodology (in percentages)						
29	Methodological choices are described	8	40	28	20	0	4
30	Methodological choices appear appropriate	0	44	20	28	4	4
31	The evaluation's limitations are reported	12	20	20	16	0	32
32	The evaluation report explains the selection of any sample	0	0	0	0	0	100
	Summary of ratings	**5**	**26**	**17**	**16**	**1**	**35**

Methodological rigour is essential in undertaking evaluations in a development context, especially as interventions often are difficult to quantify and to measure. Therefore, it is crucial that an evaluation report clearly describes the methodological choices, their appropriateness in relation to the given context, and any limitations that the evaluation might be suffering from. Overall, the 25 evaluations assessed scored unsatisfactorily in this regard. Information on 35 per cent of the criteria was missing or had not been taken into consideration. This low incidence is largely explained by the fact that no evaluation explained the selection of samples (criterion 32). Overall, the assessed evaluations do not reveal any particular trend regarding this group of criteria.

The criterion, "description of methodological choices", was rated either "very poor" or "poor" in 48 per cent of the sample (eight per cent; 40 per cent respectively). Similarly, the criterion, "appropriateness of methodological choices", was rated "poor" in 44 per cent of the sample (11 reports).

4.3.8 Assessment of evaluation findings (5 criteria)

		Very Poor	Poor	Fair	Good	Excellent	Missing
	Assessment of Evaluation Findings (in percentages)						
33	The intervention's strengths and weak-nesses have been assessed	0	4	40	44	0	12
34	Methods for assessment of results are specified and appear appropriate	12	28	12	16	4	28
35	Indicators are used as a basis for results assessment and seem appropriate	0	0	0	12	12	76
36	Evaluation findings are relevant to the intervention being evaluated and the purpose of the evaluation	0	12	20	44	20	4
37	There is a basis for the relationship between the purpose of the evaluation, methods used, data collected, and findings clearly presented in the final report.	4	20	24	48	4	0
	Summary of ratings	**3**	**13**	**19**	**33**	**8**	**24**

An assessment of the evaluation findings naturally depends upon a multitude of factors. Here, the assessment is limited to five broad criteria covering the crucial components for an effective assessment. The overall assessment in this section resulted in 41 per cent of the criteria being rated “good” or “excellent” (33 per cent; eight per cent respectively); this reflects a positive indicator for the sample quality in this aspect. Nineteen per cent of the criteria were rated “fair”, and 16 per cent were rated either “poor” or “very poor” (13 per cent; three per cent respectively). Twenty four per cent of the criteria were either missing or not applicable, across the entire sample.

The evaluation findings have to be relevant to the intervention being evaluated. 64 per cent of the sample (16 reports) were rated either “good” or “excellent” in this aspect. Similarly, in 52 per cent of the reports there was a sound basis for the relationship between the purpose of the evaluation, methods used, data collected, and findings clearly presented in the final report (13 reports rated as either “good” or “excellent”: 48 per cent; four per cent respectively). Indicators are rarely used as a basis for results assessments (they were absent in 76 per cent of the sample).

4.3.9 Assessment of the DAC criteria (5 criteria)

		Very Poor	Poor	Fair	Good	Excellent	Missing
	Assessment of DAC Evaluation Criteria (in percentages)						
38	There is an accurate assessment of efficiency	4	36	12	8	0	40
39	There is an accurate assessment of effectiveness	8	40	24	12	8	8
40	There is an accurate assessment of impact	0	36	12	4	8	40
41	There is an accurate assessment of sustainability	8	36	24	0	0	32
42	There is an accurate assessment of relevance	4	48	16	8	0	24
	Summary of ratings	**5**	**39**	**18**	**6**	**3**	**29**

The DAC evaluation criteria are internationally recognised standards that most donors and development partners have acknowledged as suitable yardsticks to be used in evaluations. The criteria of effectiveness, efficiency, relevance, impact and sustainability are constantly referred to, in both project evaluations and higher level evaluations. The 25 evaluations assessed in this present study scored poorly as assessed against the DAC standards. Overall, 44 per cent of the criteria in this section were rated as either “very poor” or “poor” across the sample reports (five per cent; 39 per cent respectively). A further 18 per cent of the criteria were rated as “fair”. Only nine per cent of the criteria were rated as either “good” or “excellent” (six per cent; three per cent respectively). Almost a third (29 per cent) of the criteria were either missing or not applicable. However, it is important to bear in mind that most evaluations do not take all five DAC criteria into account, and that this is not stipulated in the terms of reference of many evaluations.

Efficiency – the 25 evaluations contained very limited assessments of efficiency. In 40 per cent of the sample (10 reports) assessments was missing altogether. In a further 40 per cent of reports, the efficiency criterion was rated as either “very poor” or “poor” (four per cent; 36 per cent respectively). Overall, the sample reports do not provide accurate assessments regarding efficiency. This is an area that needs attention as most terms of reference require information in this regard.

Effectiveness – there was at least some effectiveness assessment in 92 per cent of the sample (23 reports). Nonetheless, the quality of these assessments varied greatly. Only 20 per cent of the sample were rated as either “good” or “excellent” (12 per cent; eight per cent respectively). Almost 50 percent of the reports were weak in this regard: 48 per cent were rated either “very poor” or “poor” (eight per cent; 40 per cent respectively). 24 per cent were rated “fair”. Considering the fact that effectiveness is an evaluation criterion that is often requested in terms of reference, this is an area that requires attention.

Impact – in recent years donors and development partners have increased their focus on evidence based methodologies in evaluation. The relative scarcity of good quality impact evaluations has been addressed by various institutions.[34] The sample of evaluations assessed in this evaluation confirms some of the problems associated with conducting good quality impact evaluations. None of the evaluations assessed is a genuine impact evaluation. However, there was some kind of assessment that can be attributed to an impact feature in 60 per cent of the sample (15 reports). Twelve per cent (three reports) were regarded as either "good" or "excellent" in this aspect. A further 12 per cent were rated as "fair". Thirty six per cent (9 reports) were rated "poor".

Sustainability – there was some form of sustainability assessment in 68 per cent of the sample (17 reports). In general, the quality of these assessments was considered poor. 44 per cent of the sample (11 reports) were rated either "very poor" or "poor" (8 per cent; 36 per cent respectively). A further 24 per cent were rated "fair".

Relevance – there was some form of relevance assessment in 76 per cent of the sample (19 reports). The overall quality of these assessments was considered low. 52 per cent of the reports were rated as either "very poor" or "poor" (four per cent; 48 per cent respectively). 16 per cent were rated "fair" and only eight per cent were rated "good". Considering the fact that this information is crucial for UNDP in determining the relevance of its projects and programmes at the country level, this is an area in which attention is needed.

4.3.10 Completeness of the report (7 criteria)

		Very Poor	Poor	Fair	Good	Excellent	Missing
	Completeness / Conclusions and recommendations (in percentages)						
43	The evaluation presents conclusions, recommendations and lessons learned separately and with a logical distinction between them	8	32	24	20	16	0
44	Conclusions are focused directly on the evaluation questions and each conclusion is supported by findings and/or data	8	24	40	20	0	8
45	Conclusions are clear and consistent	8	20	32	32	4	4
46	Recommendations are relevant and targeted to the intended users	0	8	32	52	8	0
47	Recommendations are practical and may be translated into decisions	0	24	40	32	4	0
48	Lessons learned are relevant and targeted to intended users	4	12	28	40	8	8
49	Lessons learned are practical and appears useful	4	16	48	20	4	8
	Summary of ratings	**5**	**19**	**35**	**31**	**6**	**4**

[34] For example, the 3IE initiative (*http://www.3ieimpact.org*) and the NONIE network (*http://www.worldbank.org/ieg/nonie*).

The completeness of report is an essential part of the entire evaluation, as it is closely associated with the actions that are expected to come as result of the evaluation. In order to maximise the utility of an evaluation, conclusions, lessons learned and recommendations, need to formulated in a clear and unbiased manner, and be firmly anchored in the evaluation findings. The evidence collected in this evaluation is equivocal in this regard. Overall, 37 per cent of the criteria were rated “good” or “excellent” in this section across the sample reports. A further 35 per cent were rated “fair”, and 24 per cent were rated either “very poor” or “poor” (five per cent; 19 per cent respectively).

4.4 Comments on the assessment of outcome evaluations

The quality of the sub-sample of outcome evaluations (nine reports) was uneven. This relatively small sample of outcome evaluation reports could potentially have affected the accuracy of the assessment made here, but the findings do point in a certain direction which enables generalisation for drawing conclusions.

A large proportion of the assessed evaluations lacked clarity in terms of what information sources were used to underpin the analysis of outcome level results. This was further exacerbated by vague methodological descriptions and choices. Often the evaluations have coherent objectives and purposes which are consistent with UNDP’s corporate objectives. However, the analysis and the focus of the evaluations are concerned with project activities. This limits the relevance of the reports, in that they do not accurately analyse results at an outcome level.

The issue of attribution – in this case how UNDP-funded or UNDP-managed interventions have contributed to development results at the country level – is not sufficiently addressed in general. This represents a considerable flaw in the manner in which outcome evaluations are carried out. This feature should separate outcome evaluations from standardised project- or output-oriented evaluations, but this study has found this is not occurring. One potential reason why outcome evaluations have not been executed satisfactorily is that not enough resources are provided, in relation to both the scope and design of the evaluation. Furthermore, the confusion created by various contradicting strategic objectives and reporting frameworks in UNDP, in combination with vague programme objectives, has hampered the quality of outcome evaluations, and therefore also their potential as a reliable guide for decision-making. However, the transition into conducting outcome evaluations instead of output evaluations does require a real change of culture, both internally at UNDP as well as within its partners. The generally unsatisfactory results described above should take into account that this is an ongoing process, and that the transition period corresponded with the 2004-2006 period from which the sample of evaluations were selected.

5 Main findings from field visits

This section is based on interviews with UNDP country office staff and various stakeholder groups, relevant to evaluations commissioned by the country offices in Kenya, Uganda, and Nepal.

5.1 Aspects affecting UNDP country level evaluation quality

There appears to be a large pool of evaluation consultants in Kenya, while in neighbouring Uganda there appear to be fewer active evaluation consultants. In Nepal, there appears to be a real dearth of active evaluation consultants. All evaluation consultants interviewed, both international and national consultants, had several years of professional experience in evaluation. Most had worked with a variety of different actors in the development community: government, bilateral donors, NGOs, or multilateral organisations. All had received some type of evaluation training, and most had a background in social research or academia. However, none had received any specific training in relation to outcome evaluation methodology.

The UNDP evaluation policy is clear on the importance of consultants being independent from the programme or project that is being evaluated. There is a trade-off between hiring an outsider to conduct an evaluation. The evaluation consultant provides other perspectives and is less likely to be biased, however they might lack contextual understanding.

This problem could be exacerbated by the limited time set aside for background studies. All consultants mentioned time constraints as a serious obstacle for conducting objective, credible and relevant evaluations. This issue has been raised by consultants with UNDP on several occasions, however no change has yet occurred. This limits the ability of consultants to develop a thorough understanding of the programme to be evaluated. It might also reduce the length of time available in the field. Also, document analysis and reviews of background information may become cursory and hurried.

Some consultants expressed a frustration at not being sufficiently involved in follow-up activities of evaluations they have conducted, such as workshops and seminars. This issues could be seen alternatively as a waste of valuable knowledge, or as a way to maintain independence and to avoid the risk of potential conflicts of interest. The consultants expressed divergent views in relation to the ability for them to work independently. Some found UNDP to be very supportive in terms of providing necessary assistance and respecting the independence of the consultants' work, while others felt constrained by the organisation's requirements (for instance, some consultants were required to be present at UNDP's office when drafting their report). Some felt UNDP was reluctant to share relevant information sources. The fact that issues related to travelling during the evaluation mission are handled by UNDP was also perceived differently. Some appreciated this as a form of assistance, while others perceived it as a constraint.

All consultants interviewed reported heavy UNDP bureaucracy as an obstacle, implying prolonged recruitment processes, late reimbursements and a lack of transparency in the recruitment process when setting up evaluation teams. However, the CO in Kenya contradicted this view, claiming that the recruitment process is competitive. A number of consultants testified that UNDP had expressed dissatisfaction about the consultants' evaluation recommendations on the basis that they were "too critical".

Payment from UNDP was generally seen as insufficient by the national consultants (INGOs and other donors were reported to offer substantially larger remuneration, sometimes double that of UNDP). Consequently, the best qualified (national) consultants may not consider consultancies for UNDP. Meanwhile, UNDP is well regarded in the donor community: working with UNDP is regarded as a rewarding experience: "… you sometimes accept very low payment to get the experience". According to interviewees at the country offices, UNDP is aware of the problem of low payments and is in the process of reviewing reimbursement levels.

5.2 Conduct and use of evaluations

5.2.1 In-house UNDP evaluation capacity

The evaluation capacity of the three country offices visited in this study varied. The Kenyan CO has an M&E analyst with the core responsibility for all evaluation activities of the office. The CO in Uganda has, until the beginning of 2008, only had an M&E focal point, which is simply an "added on" task for a staff member who has other responsibilities. The office has not established a clear mandate for a dedicated M&E analyst, nor has it appropriated sufficient financial and human resources for such a position. Since 2004, UNDP's Nepal CO has developed a stand-alone M&E unit with two staff working full-time on issues related to M&E. The M&E unit is responsible for: i) providing strategic advice to senior management based on analysis, ii) strengthening monitoring, iii) managing outcome and project evaluations, iv) provide advisory services and capacity building in RBM, M&E and reporting, and v) knowledge management and strategic partnership.

Box 4: Example from Nepal CO: The M&E unit of the CO in Nepal appears very active and motivated. A number of country-adapted guidelines have been developed in the field of M&E (such as "Guidelines for evaluation – UNDP") that provide complementary information to the Evaluation Policy of UNDP (2006) and the UNDP Handbook on Monitoring and Evaluation for Results, in relation to the practical aspects of conducting an evaluation in Nepal. The guidelines also address issues such as national ownership and alignment with national priorities, quality, capacity building, dissemination and use of evaluation. A user-friendly "Monitoring and Evaluation Framework – UNDP Nepal" has also been developed, addressing issues of field monitoring, with a template for field project monitoring reports.

Consultants interviewed in both Kenya and Uganda expressed the view that the COs do not possess strong enough internal evaluation capacities and are hence not able to provide the necessary support. Consultants generally considered there was little knowledge and awareness within COs of both the purpose of, and methodologies associated with, outcome evaluation.

Several interviewees at the COs considered that guidelines and methodologies for evaluation were both adequate and concrete. However, a need for further training was expressed, in order that the guidelines and methodologies could be applied professionally. The COs described their relationship with UNDP's headquarters and EO as generally good. However, the COs considered the EO's understanding of the implication of the country context for evaluation to be limited.

In order to improve the ability of programmes to be evaluated, the respondents commonly expressed the view that that the M&E function of a CO must be an integral part of programming. When M&E is isolated and is poorly planned, the utility and relevance of evaluations is reduced. In most cases in COs examined in this study, there was no separate budget for evaluation. A separate evaluation budget was generally seen by respondents as a prerequisite for reinforcing the role and quality of evaluations.

A general reluctance to work towards outcomes instead of outputs could be noted throughout all of the interviews with the various categories of respondents, in particular internally at UNDP COs. Programme officers tended to prefer measuring their particular activity and indicated only a limited interest in the activity's contribution to the achievement of an overarching outcome. Individual programme managers indicated that there are not always clear incentives in place that stimulate thinking in terms of outcomes rather than outputs. Outputs are less abstract and generally easier to grasp. Overall, amongst all respondents, there appeared to be a low level of acceptance of outcome evaluations, in terms of purposes, utility and methodological issues.

Box 5: Example from Kenya CO: The central drive to change UNDP into a more results-orientated organisation has, at the Kenya CO, resulted in a radical overhaul of the country office organisation. This has also been partly precipitated by a substantial increase in funding, from USD 9 million in 2003 to USD 22 million in 2007, an increase of 144 per cent. With the country representative as the main driving force, an ambitious mainstreaming strategy was devised and followed through the period 2005-2007, while at the same time necessary adjustments in the structure of the office were made. As the approach to the project cycle became more holistic within the organisation, business service centres were set up to serve the office, as well as a programme unit as a planning apex.

This has resulted in each unit now setting their own targets, which are reported quarterly and followed up annually. Each staff member of the office has been certified in the UNDP results based system, and can now log on and report through an internet-based reporting system, spreading the ownership and participation of the results reporting throughout the whole office.

This massive organisational change was primarily conducted by mapping out and using relevant capacity among the existing staff of the country office, rather than by using external consultants. This has led to a better knowledge, and higher use, of the potential of the staff, which has boosted the morale and involvement of the office staff, with clear tangible effects. The success of this process has led to the model spreading horizontally to other country offices in the region.

In relation to the specific human rights and democratisation partnership currently in place in Kenya, UNDP has "inherited" partner NGOs from previous donors. Usually an annual capacity assessment is made by external evaluators commissioned by UNDP, in order to strengthen performance and partnerships.

5.2.2 Outcome evaluations – concept and methodology

Respondents within all three COs indicated the core importance of the motivation and interest of CO senior management for a successful transfer to using outcome evaluations as a tool for improved programming and for strengthening the role and use of evaluations. Since UNDP does not have a corporate policy dictating an explicit amount of resources that should be allocated to evaluation at the COs, the responsibility of promoting M&E issues falls on the senior management at each CO. In the case of Nepal, senior management has taken the initiative, with the establishment of an M&E unit with two staff working fulltime on issues related to M&E.

All consultants reported a clear mismatch between the planning of activities and evaluations. One respondent indicated: "Programs have not been designed for measuring outcomes". A reported challenge was that outcomes are being differently defined in different strategic documents which sometimes overlap. This creates uncertainty about what is actually to be achieved, and it makes accurate assessments a very challenging task. The problems associated with the shifting strategic documents were addressed by all CO respondents interviewed. Moreover, the rapid shift of strategic documents also poses problems for partners (government, other donors and implementing partners). However, there was a common understanding among respondents at the COs that the corporate performance systems needed to be revised when they are neither working nor delivering on their objectives.

A general problem, according to the consultants interviewed, is that project activities have not been linked to specific outcomes. Instead, executed projects are "squeezed in" to fit pre-determined outcomes. This results in a "backwards way of working". The process of formulating achievable and realistic outcomes has proven problematic. Some consultants claimed that, in reality, outputs – which are either inadvertently or deliberately mis-labelled as "outcomes" – are assessed. Misunderstandings about what is to be regarded as an "outcome" frequently occur, which further complicates the possibility of conducting good quality outcome evaluations. Another challenge that affects the ability of programmes to be evaluated is the frequent absence of baselines against which to assess outcomes. Sound assessment of the achievement of an outcome requires that a baseline be in place from the beginning. In certain cases baselines had to be created subsequently in order to evaluate the given project or programme.

There also appeared to be a general consensus among stakeholders that the outcomes stipulated in relevant strategic documents have been overly optimistic or too far-reaching. Realistic achievement of these outcomes first requires a substantive change in culture and systems. The respondents claimed that this is not achievable within a programme period of only a few years, and further, that an accurate assessment of outcomes requires that the assessment take into account a longer time period. However, this perspective is problematic considering the current process within UNDP, where performance indicators and policy documents change frequently.

According to respondents from the Nepal CO, challenges have also been faced in defining outcomes that are owned by other partners. The different strategies (UN common, UNDP corporate, UNDP country specific) need to be more closely linked to one another and, in particular, to the partner country's priorities.

In Uganda, according to evaluation consultants, the main challenges involved in the transition from project evaluations into outcome evaluations consist in: i) aligning evaluations and evaluation processes with government priorities, ii) the programme design – the ability of programmes to be evaluated, iii) training needs and the need for both a change of culture regarding outcome evaluations, and for raising awareness and knowledge – internally and externally – about outcome evaluations, and iv) difficulties in setting up appropriate outcome indicators in certain programmes.

CO staff perceived project evaluations as a valuable tool for follow-up and improving UNDP's performance. In general, project evaluations were perceived as "much more practical" than outcome evaluations. Project evaluations are no longer mandatory, except when stipulated in a co-financing donor's protocol. However, interviews supported the view that it is understandable why UNDP still relies on project evaluations, since the bulk of the organisation's workload at the country level is rooted in project activities. In this respect, project evaluations are a useful tool for determining the continuation and/or re-direction of a project.

5.2.3 Evaluation partnerships

Several interviewees noted the problem of non-inclusion: the link between the CPAP and the partnership with different NGOs and implementing partners has not been clear; the implementing partners having very little input into the CPAP process. Several implementing partners, in particular NGOs, regarded the dialogue in relation to the CPAP as "a very vertical communication" between UNDP and the government.

COs, as well as government representatives and NGOs, reported evaluation at UNDP still being a "very internal thing", conducted for accountability reasons. As reported by the Uganda CO, the first two outcome evaluations conducted by the CO (in 2005) were a compliance exercise rather than demand driven: "UNDP headquarters pushed for it getting done." This point of view is shared by many interviewees; some government representatives claimed that all evaluations are essentially being conducted because donors require it.

It is evident from the interviews with evaluation stakeholders that stakeholders are not involved to a great extent in the evaluation process. Generally, UNDP was said to be "just verifying things: not really taking in our conceptions and understandings." Several evaluation stakeholders reported that UNDP is not interacting sufficiently during the evaluation process.

Several evaluation stakeholders interviewed reported that there is generally little coordination between UNDP and other donors, resulting in much overlap and duplication. There are examples in which several donors have undertaken evaluations of similar programmes or sectors without coordinating their activities. Testimony suggested that more collaboration between partners that operate within the same sector should be pursued. In general, UNDP appears to rarely use other donors' evaluation reports.

5.2.4 Evaluation use

It has proven problematic to ensure that evaluation results are sufficiently used and adequately taken into consideration in new projects and programmes. Finding the proper tools for a high degree of evaluation use is not an easy task. A clear method to ensure benefits accrue from the results of the evaluation for the next planning period appears to be missing in most countries visited. The original reason for conducting evaluations sometimes appears to be neglected or disregarded, in that they are not sufficiently used.

Some consultants interviewed claimed that evaluations in general provide sparse information on improvement possibilities and recommendations are not sufficiently clear. Meanwhile, several interviewees expressed the view that evaluations in general tend to confirm what is already known, rather than bring new knowledge; hence they do not add value.

Staff interviewees at all three COs mentioned the establishment of the UNDP management response system as an important measure for improving the general use of evaluations. The integration of the system in order for it to become fully effective remains an ongoing process. The system has two parts: 1) presentation of management response to key issues and recommendations and key follow-up actions, and 2) follow-up by tracking actions. The process of preparing a response provides an opportunity to engage in a dialogue with relevant stakeholders to reflect on the evaluation process, findings, recommendations and lessons.

Further, the ERC database was viewed as an important instrument for improved future evaluation use, permitting transparency of evaluations and evaluation use. At the present time, the ERC remains under development. The corporate guidelines for the ERC and its role in evaluation accountability are clearly stipulated. Some interviewees expressed that a better steering (with clearly stated responsibilities) by UNDP headquarters of how the ERC is managed would strengthen the role of the ERC. Hence, there appears to be some problems associated with the manner in which the use of the ERC has been communicated from headquarters to the COs, and how its guidelines have been disseminated. To date, it is the responsibility of each CO to update the ERC with evaluation plans and ongoing evaluation activities.

Several government representatives interviewed reported that an important drawback for a wider use of evaluations is that UNDP is not sufficiently communicating the results of its evaluations. Some implementing agencies claimed that they only received some “drivers” in relation to issues to improve but were not informed about the evaluation and evaluation process that arrived at these conclusions and recommendations.

Box 6: Example from the Nepal CO:

The country office in Nepal has made concrete efforts to ensure that evaluations are considered and taken into account into programme and planning processes. The efforts listed below can at least partly be attributed to the existence of the M&E unit, which is specifically responsible for issues related to M&E.

- Before beginning the preparation of the UNDAF/CPD/CPAP in early 2007, an in-house review was conducted of all evaluations, reviews and studies that had been conducted between 2003 and the end of 2006, and a consolidated report was prepared highlighting recurring findings and cross-cutting issues that the new programme needed to address. The report was used as a reference throughout the CPD/CPAP preparation process, and provided the basis for the "lessons learned" sections of these documents.
- When a new project document is prepared, the M&E unit, among others, must review and sign-off on the new document. Among other things (such as checking that the UNDAF and CPAP outcomes and outputs are correctly reflected, that the M&E sections are complete and the necessary annexes have been included), the M&E unit checks whether an evaluation of a previous phase of the project has been conducted, and assesses whether or not the new project reflects the recommendations and learning from that prior evaluation.

All evaluations, in addition to being linked to the ERC, are linked to the in-house "Comp@ss" system (Country Office Management of Programmes System), so that staff can easily refer to them.

5.3 The strengthening of national evaluation capacities

Several of the evaluation stakeholders claimed that UNDP was more concerned with the accountability of its own programmes rather than supporting changes on the national level that could strengthen the role of M&E throughout government. Increasing internal resources dedicated to evaluation was generally seen by interviewees as a step in the right direction, but not likely to bring about substantial change. Interviewees suggested that for real change to occur, UNDP's programmatic direction should be altered towards making a commitment in support of capacity development in evaluation in partner governments.

Of the three COs visited, the Kenya CO appeared to have collaborated the most with the partner country in terms of national evaluation capacity strengthening. Here, the support has consisted of funding an evaluation expert to work within the government in order to strengthen the M&E function. No such support appeared to have been given in Uganda and Nepal. In Nepal, few activities had been undertaken in this area. The clear reason for this was the complex country context, rather than a lack of interest or motivation on behalf of the CO.

Government representatives in Uganda reported that UNDP's mandate as a capacity-building agency in terms of M&E has not been fulfilled. UNDP Uganda has, until now, been working outside the national M&E system, as provided by the Ugandan government through the Poverty Eradication Action Plan (PEAP). The M&E activities of UNDP Uganda are not aligned with the national M&E framework, and hence do not support the long-term development of M&E in the context of Uganda.

One consultant in Nepal identified the tension between donors and the government as an obstacle to sustainable capacity building. Another aspect of this problem is that a considerable portion of the budget of any given donor-funded project will go to technical experts in cases where these are part of the project staff. This may not always be an effective use of funds. Up to 60 per cent of the budget of a project can be spent on salaries for "technical advisors" that have been internationally recruited to support the project. National capacities should, it was argued, be utilised to a greater extent. The differences in wages appears to deplete the morale and motivation of local staff, which impedes national capacity building. In cases where national capacities exist, these are currently not being used optimally.

In relation to the issue of ownership of evaluations, government representatives in all three countries claimed that donors in general have specific (hidden) objectives in financing certain activities, and are therefore not inclined towards the government taking the lead in evaluation (and in therefore becoming obliged to adapt to the government's recommendations). The general consensus appeared to be that donors tend to retain a certain degree of influence over the results of evaluations. The explanation commonly given for this was that governments possess insufficient evaluation skills and resources and that there is a reluctance to actively participate in evaluation processes.

Including a government representative as a team member in a UNDP-led evaluation appears to have strong potential. This is a capacity development measure with possible gains for both UNDP and government. However, a government representative interviewed in Nepal reported having little influence in the evaluation process. Also, no reimbursement for participating was provided to the government representative. Interviews with CO representatives in Kenya and Uganda did not suggest that these offices had been involved in including government representatives in evaluation teams. The reimbursement issue can be regarded in either of two ways: it is important to create the proper incentives in order to maximise the participation of government representatives; however, at the same time, reimbursements to national partners for participation in evaluations may be seen as contradictory to the principles of the Paris Declaration.

6 Conclusions

Based on the findings in the chapters above, the following conclusions may be drawn in relation to: i) the quality of evaluations commissioned by UNDP's country offices, ii) the conduct and use of these evaluations, and iii) UNDP's efforts to strengthen national evaluation capacities. Some concluding remarks are provided at the end of this chapter, summarising issues that require particular attention to facilitate a more results-orientated Swedish involvement in multilateral development cooperation in general, and with UNDP in particular.

6.1 Quality of evaluations commissioned by UNDP's country offices – aspects affecting evaluation quality

1 In assessing the quality of a sample of evaluations commissioned by UNDP's Country Offices, only six per cent of the criteria were rated "excellent", whereas the ratings were evenly spread between "good", "fair", and "poor" and "very poor" (25 per cent, 20 per cent and 24 per cent respectively). It can be concluded that the overall country level evaluation quality of UNDP is poor.

2 The quality assessment indicated the average time spent on evaluation was two to three weeks. As to costs of evaluation, little information was provided in the assessed sample (Chapter 4.1). All consultants who were interviewed indicated that the time allocated for conducting evaluations is not sufficient (Chapter 5.1). This hinders a thoroughly assessment of relevant information, which may have serious implications for the quality of the evaluation and the reliability of the conclusions and recommendations. Limited resources for evaluation was repeatedly raised as a constraint for carrying out good quality evaluations at the country level. Resource requirements appear not to be identified at an early stage in the programming cycle. Preferably, resource requirements should be addressed at the design phase. Also, cost-sharing arrangements with other organisations or government could be pursued to a much greater extent. The UNDP Evaluation Policy states that senior management should "ensure adequate resources for evaluation."[35] Resources allocation for evaluation, both human and financial, appears to be organised haphazardly and lack coherence.

3 As to the overall structure of the evaluation reports, the majority of the reports proved relatively well structured in terms of clarity and directness (Chapter 4.3.1). This, together with well defined purpose, objective and scope of the evaluation (Chapter 4.3.2), is essential for an appropriate and efficient use of the evaluation. However, the absence of executive summaries as well as annexes to which the report refers – which was the case for several reports in the sample – does constitute a serious hindrance to an efficient use of an evaluation report.

[35] UNDP (2006): Supra note 24, §20.

4 In half of the evaluations assessed, the information on evaluator competencies was rated as either “poor”, “very poor”, or this information was missing (Chapter 4.3.3). However, the information collected through interviews gave a different picture, indicating evaluator competencies were generally high, although generally lacking outcome evaluation methodology training (Chapter 5.1).

5 In relation to information on independency in reporting, evaluation ethics and threats to reliability and validity (Chapter 4.3.3), this was generally of very low quality, or the information was entirely absent in the evaluations assessed.

When the independence questions were raised in interviews with consultants, responses indicated that the UNDP is too closely involved in the evaluation arrangements, which could affect the independence of the evaluation results. Non-transparent recruitment processes and pressure felt by consultants against providing conclusions that are too negative were also raised at this issue. In light of these serious concerns, it is hence not clear how COs ensure independence in the evaluation process. It is important that CO senior management is aware of the independent nature of evaluation while providing adequate support to an evaluation unit or officer. The same obviously applies to issues of evaluation ethics and threats to reliability and validity in reporting. As mentioned above, time constraints imply a serious constraint for conducting reliable evaluations with valid conclusions and recommendations.

6 The criteria assessing the provision of information sources (Chapter 4.3.4) used in the evaluation received low ratings in the quality assessment. Sources of information were described unsatisfactorily, while limitations of the information used were rarely discussed at all. These lacks impede an independent assessment by the reader of the appropriateness of information used for the evaluation at hand. Information should be strengthened in this aspect.

7 Evaluations should include a description of the intervention being evaluated, its specific features and components (Chapter 4.3.5). The sample scored satisfactorily in this aspect, permitting an assessment of the accuracy of the evaluation analysis.

8 From the quality assessment it is possible to conclude that the DAC evaluation criteria (Chapter 4.3.9) are not being applied systematically or coherently. Although constantly referred to in Terms of Reference, they are not being correctly applied. The sample of evaluations assessed scored poorly, assessed against the DAC standards.

9 The quality and utility of outcome evaluations was perceived as poor among interviewees. In particular, interviewees noted that outcome evaluations failed to answer pertinent questions and that they lacked a rigorous assessment of the contribution to the outcome(s) in question. The credibility of outcome evaluations has suffered from weak monitoring and the absence of reliable data. More time and resources is needed in planning and managing outcome evaluations.

6.2 Conduct and use of evaluations

6.2.1 In-house UNDP evaluation capacity

10 The in-house evaluation capabilities and arrangements appeared to vary between the COs visited (Chapter 5.2.1). In general, the role and responsibilities of the M&E focal point appears vague and varies between offices. The institutionalisation of an M&E officer position (with the overall responsibility of the CO's evaluation activities) appears to strengthen the importance and quality of evaluation activities at the CO. However, the interest paid by senior management at the COs is critically important. Since UNDP does not have a corporate policy dictating explicitly what resources should be allocated to evaluation, the responsibility of raising M&E issues falls on the senior management at each CO.

11 Of great importance is the evident need for a change of culture. Evaluation should be recognised as a tool for improved effectiveness rather than as a compliance exercise. An important step for reinforcing the role and quality of evaluations would be to further integrate evaluation into the overall programme planning cycle.

12 In-house evaluation capacities need to be strengthened, in terms of both knowledge and human and financial resources. There appears to be a need for greater support for managing for outcomes at COs. To promote effective results reporting, senior managers at COs should be held accountable for demonstrating results.

6.2.2 Outcome evaluation – Concept and methodology

13 The rapidly shifting and sometimes overlapping strategic policy documents appear to pose severe threats to the planning and conduct of evaluations (Chapter 5.2.2). Moreover, the corporate goals and service lines set by UNDP headquarters have proved too numerous and too broadly defined, offering too much room for interpretation. This has led to COs manipulating their programmes to fit pre-determined outcomes. The current process favours a "backwards way of working", in which outputs are being subsequently linked to outcomes (Chapter 5.2.2). With a wide range of strategic – sometimes overlapping – documents stipulating different outcomes to be achieved, it is difficult to see how the results of outcome evaluations feed into future activities.

14 The shift towards a system in which evaluations measures outcomes instead of outputs is posing challenges for UNDP. Outcome evaluations have to date not had the impact that they were intended to have. Far-reaching and unrealistic outcomes, combined with a lack of knowledge about outcome evaluation methodologies, have been identified as obstacles (Chapter 5.2.2). Related to this is the time perspective. For a useful assessment of sustainable outcomes, a longer perspective is needed. However, this is problematic in view of UNDP's current planning, where performance indicators and policy documents change frequently. The fact that outcomes (in most cases) can only be assessed several years after the termination of a programme requires a change in current programming. To maximise the results of an outcome evaluation, the evaluation

should be planned to be conducted several years after termination of the programme.

15 There is an apparent problem in the manner in which performance indicators are used. Indicators are not always being defined at the programme design phase, which makes subsequent monitoring and evaluation activities problematic. These problems are attributable, to a certain extent, to a lack of appropriate baselines against which to measure (Chapter 5.2.2). However, informants from UNDP Kenya indicated that efforts are currently underway to ensure that indicators and baselines are defined at the design phase of any given project or programme.

6.2.3 Evaluation partnerships

16 The quality assessment of evaluations indicated the involvement of stakeholders as either "fair" or "good" in 76 per cent of the sample evaluations (19 reports) (Chapter 4.3.3). In direct contradiction to this finding, interviews with implementing partners indicated the involvement of stakeholders in the CPAP process as generally weak, and in strong need of improvement (Chapter 4.3.4). Evaluation stakeholders appear not to participate sufficiently in the UNDP evaluation process. The collection of these groups' conceptions and understandings appears to be weak. The contradictory findings on this issue may be explained by varying conceptions of "involvement of stakeholders". This is an area which requires improvement, as the utility and credibility of evaluations directly depends upon the extent to which stakeholders are involved. The buy-in of stakeholders is essential for the evaluation to be used constructively.

17 There appeared to be a general lack of national ownership of evaluation. Evaluation at the country level is driven by donor demands, with little involvement of national capacities (Chapter 5.3). It was evident in the three countries visited that demand for evaluation is relatively weak. This may be explained by the absence of, or weak, systems for accountability of public policy at the national level. Among the countries visited, Kenya appeared to have the strongest national systems for evaluation, while Nepal lagged most in this regard. A change in culture towards an improved understanding of the role and use of evaluation, and making evaluation more relevant to public interests, would strengthen national ownership.

18 Donors appear reluctant to cede control to the government in leading evaluations (Chapter 5.3). Explaining and accepting this through a lack of resources and government interest does not appear to be an acceptable long-term position, particularly in light of the effect of international commitments such as the Paris Declaration.

19 Interviews suggested that donors often tend to have specific objectives in evaluation, that are not necessarily consistent with the government's priorities. These issues need to be examined, otherwise this issue will remain an obstacle to both an improved alignment between donors' and government interests, and strengthened national ownership. Furthermore, efforts need to be made to

ensure that outcomes are owned by the country. In particular the CPAP should be more closely linked to the partner country's priorities.

20 A factor that should clearly be taken into account when determining the extent to which alignment is appropriate is the country context in which UNDP operates. In countries with limited institutional capacity, complete alignment is not practicable, however a continuous dialogue with the government on these issues is an appropriate alternative.

21 There appears to be little coordination and collaboration between UNDP and other donors in evaluation work, resulting in much overlaps and duplications of work.

6.2.4 Evaluation use

22 The quality assessment indicated that evaluations appear generally relevant to the intervention being evaluated (Chapter 4.3.6). In approximately half of the evaluation sample the relationship between the purpose of evaluation, methods used, data collected and findings was clearly presented (Chapter 4.3.8). For an effective use of evaluations, it is key that the evaluation appears relevant to the intervention being evaluated, and that the links between different aspects of the evaluation are clear. This appears to be an area for improvement.

23 Completeness of the evaluation report (Chapter 4.3.10) is evidenced by conclusions, lessons learned and recommendation being clearly presented and firmly anchored in the evaluation findings. The results of the quality assessment were inconclusive as the ratings were evenly spread across the scoring scale. Interviews indicated that evaluations generally provided sparse information in relation to improvement possibilities and recommendations were not sufficiently clear. This was confirmed by the quality assessment, in which the criteria concerning "recommendations being practical and may be translated into decisions" scored a low rating.

24 Partners' use of UNDP evaluations appeared generally limited, which was in some cases explained by an insufficient understanding of the role and use of evaluation results, and in other cases due to failure of the CO to share such results (Chapter 5.2.4). For improving UNDP partners' use of evaluations, results need to be clearly communicated. Also, the involvement of different stakeholders groups during the evaluation process – providing these groups the possibility to influence evaluation design and questions – is necessary in order to achieve a broader utility of evaluation results.

25 Interviews with CO representatives indicated that the internal use of evaluations was not coordinated or carried out in a systematic manner. A clear method for utilising evaluation findings and for feeding these into new programming appears to be absent. Informants from the COs under the study indicated some uncertainty in relation to how evaluation findings are used.

26 CO representatives regard the Management Response System provided by UNDP EO in the ERC as having great potential (Chapter 5.2.4). To date, the system has not been fully institutionalised and needs to be further developed in

order to operate effectively. However, clear incentives appear to be required to ensure that COs utilise the system constructively. Some interviewees indicated that better steering by UNDP EO would strengthen the role of the ERC. To date, there are no sanctions in the event that COs do not follow up recommendations found acceptable.

6.3 The strengthening of national evaluation capacities

27 The inclusion of a government representative in the evaluation process should be seen as a measure for strengthening national evaluation capacity, and could, in the long-term, improve national ownership. However, it appears that the role of such a government representative remains relatively limited, and such representatives have little influence on the strategic selection of evaluation topics and design of evaluation. This obviously impedes an improved understanding and use of evaluation results outside of the organisation. This measure has strong potential for ensuring a wider ownership of evaluation, and should be further developed in terms of providing better opportunities for government influence in the evaluation process.

28 The availability of national evaluation consultants with appropriate thematic and methodological background varies between the three countries visited (Chapter 5.1). There is a need to further explore the possibilities of relying more on national consultants, than on international consultants with limited contextual knowledge, when undertaking evaluations. In order to strengthen national capacity building in evaluation, the focus should be in providing opportunities for evaluators from the country or within the region.

29 Allowing for increased leadership on the government side when evaluating donor financed programmes could be further explored by UNDP. Where M&E tools and methodologies are developed, and capable staff exist within government, the skills of these persons could be utilised. This is a matter of ownership, and represents an opportunity to create sustainable M&E activities in the longer-term. It is an area in which UNDP should improve in order to deliver according to its evaluation policy.

30 The UNDP evaluation policy is clear in that evaluation should always – when feasible – be conducted through national systems. In cases such as Uganda this is not being done. UNDP evaluation appears more concerned with organisational accountability than with the larger issues of increased national ownership and of capacity development, as stipulated in the Evaluation Policy. Leadership from senior management at the COs is crucial in promoting this principle.

6.4 Concluding remarks

Evaluations commissioned by UNDP's country offices constitute important building blocks in other, overarching, evaluations produced by the organisation. It is hence critically important that these are of an acceptable quality. This is not presently the case. In order to effectively manage for results – and to clearly demonstrate what results have been obtained – there appears to be a need for a strengthened framework around country office evaluation activities. This includes strengthening methodological knowledge, making evaluation an integral part of the programming cycle and creating stronger in-house incentives for conducting reliable and useful evaluations.

Unless UNDP takes serious steps to reform its evaluation system, this picture will not improve. To date, evaluations commissioned by UNDP's country offices are not subject to any formal external quality assurance. A system that checks for quality assurance is necessary to ensure that results entering the performance measurement system meet at least minimum quality standards.

With the generally strengthened results-focus within development cooperation, donors allocating financing to UNDP should endeavour to ensure that the role of evaluation within the organisation is secured and that evaluation influences how programmes are designed and implemented.

There appears to be strong potential for improvement in the area of evaluation use within UNDP. A review of strategic documents, the outcomes they stipulate, how these are formulated, and within what time frames, appears to be strongly needed. Outcome evaluations have, to date, not had the impact that they are intended to have. It is not clear the extent to which and how these evaluations are feeding into further activities. The reason for undertaking outcome evaluations – assessing performance in terms of development results – appears to be being neglected. In order to properly use the reporting of UNDP as performance evidence, a donor such as Sweden should demand an improved use of evaluations within the organisation.

Finally, the strengthening of national evaluation capacities is crucial for an increased ownership of development activities undertaken in the partner country. UNDP has, like Sweden, committed itself to the principles of the Paris Declaration. Appropriate implementation of these undertakings requires that both UNDP and Sweden ensure that these commitments are respected and that therefore appropriate measures are undertaken within UNDP for strengthening national capacities in terms of evaluation.

7 Recommendations

Based on the findings and conclusions in previous chapters, the following recommendations are focused towards improving the coherence of, and promoting a results-orientated approach towards, Sweden's multilateral development cooperation.

7.1 Measures for improving Sweden's assessment of multilateral organisations' performance

Since evaluation is an important element for improved performance information, Sweden's approach to assessing multilateral organisations' performance, in terms of relevance and effectiveness, could be further developed by including the following:

- **An "in-house evaluation capacities strengthening" check:** With an intensified focus on evaluation and results-reporting, evaluation capacities at country offices need to be strengthened. Therefore, it appears appropriate to check current evaluation capacities – at UNDP headquarters and at country offices – in relation to how these are organised and what measures have been taken for strengthening them. Strengthening evaluation capacities involves real financial costs. Hence it is important to ensure that evaluation is budgeted for in all types of activities. Ex post evaluations must be factored in as an actual cost for the project or programme; this aspect could be placed in a special reserve fund which outlives the project. The capacity check should hence include how evaluations are budgeted for.
- **A "country office senior management accountability" check:** Leadership from senior management at the country offices is crucial for improved results reporting and a strengthened focus on evaluation. It is arguable that some form of accountability mechanism should be in place to ensure that senior management of country offices fulfil their responsibilities in terms of evaluations.
- **A "capacity development and advocacy for evaluation" check:** Considering the framework of the Paris Declaration, which emphasises improved partner country ownership, it is logical to check the extent to which, and how, efforts have been undertaken to strengthen national evaluation capacities. In connection with this, checking efforts undertaken for advocating the role and use of evaluations appears appropriate. This measure might be of special interest for an organisation like UNDP, having the explicit mandate to work through national systems. Also, the information extracted through this check would provide valuable information when determining resource allocations (ear-marked funding) for this kind of activity.

- **A "partnerships in evaluation" check**: Sweden's assessments should attempt to frame the extent to which, and how, consistent with the Paris Declaration, the multilateral organisation seeks partnerships in evaluation with:

    - implementing partners,
    - other donors, and
    - the partner country

    Partnership in evaluation with *implementing partners* provides a measure of the utility of the evaluation. If implementing partners are involved in the evaluation process, it is more likely that the results will be effectively used in further activities.

    Partnership in evaluation with *other donors* entails the organisation endeavouring to work in harmony with other donors. This would include, for instance, developing a jointly agreed evaluation terminology and common formats for reporting.

    Partnership in evaluation with the *partner country* is key for an appropriate understanding of both the extent to which the organisation engages in a dialogue in relation to the achievement of common objectives, and the extent to which national evaluation capacities are used and developed. The inclusion of partner country government representatives in evaluation teams should provide better opportunities for the latter to influence the overall evaluation process, including evaluation questions and design, and hence should increase the utility of evaluation results.

- **A "reality check"**: Sweden's assessments should be complemented by a quality assessment of a limited random sample of outcome evaluations produced by the organisation. It is proposed that a simplified version of the assessment instrument used in this evaluation, covering 10 different evaluation quality aspects, should be applied in this respect.

    This assessment would provide valuable insights into evaluation quality, and permit an independent view of the credibility, utility and relevance of evaluations produced at the country level. The results of such an assessment could be raised in policy dialogue with the organisation and would, in the long-term, create further incentives for the organisation to improve the quality of evaluations.

7.2 Actions for a more coherent and strategic Swedish positioning on UNDP's Executive Board

Sweden's Strategy for Multilateral Development Cooperation (the Strategy) emphasises relevance and effectiveness in multilateral cooperation. This places heavy demands on follow-up and evaluation. To build on the results of Sweden's assessment of UNDP, it appears appropriate that Sweden is active in demanding improvements on the following issues:

- **Advocate a simplified results reporting system:** The current multitude of strategic objectives (MDGs, UNDAF outcomes, UNDP strategic outcomes, UNDP country level outcomes and country national plans) appears to hinder effective results-reporting. To reinforce the decentralised nature of UNDP and assist it to exploit its comparative advantage to a greater extent, it appears reasonable to argue for a results framework that distinguishes more clearly between corporate goals and country programme outcomes. Outcomes and indicators, directly linked to the UNDAF outcomes and national development outcomes, should be set at the country level. Through this, the effectiveness of UNDP would be assessed for each individual country rather than for the organisation as a whole. This would provide greater incentives for senior management to achieve – and report on – agreed outcomes.

  In terms of evaluations commissioned by UNDP's country offices, this simplification would advance the possibility of outcome evaluations becoming strategic instruments for decision-making. The current "backwards way of working" – where project activities are manipulated to fit overarching outcome objectives – needs to end.

  In connection with this, the longer time frames commonly required for an efficient assessment of sustainable outcomes should be taken into consideration when determining appropriate planning and reporting periods.

- **Advocate in-house evaluation capacity strengthening:** See above.

- **Advocate holding country office senior management accountable:** See above.

- **Propose M&E capacity building as a UNDP programmatic activity:** Considering the evaluation policy of UNDP, and UNDP's mandate and comparative advantage, it could be argued that providing support to partner governments in the area of M&E capacity development should be a programmatic activity of UNDP. For Sweden to advocate such a change would be consistent with the Swedish Strategy, which emphasises the importance of stimulating and developing the capacity of partner countries to perform evaluations of development initiatives.

- **Propose evaluation advocacy as a UNDP programmatic activity:** Partner countries often display a limited interest in and understanding of the role and use of evaluation. The opportunities for partner countries to actively engage in evaluation activities and to claim ownership are therefore limited. Strengthening the demand for evaluation is hence a prerequisite for any sustainable evaluation ownership and development of national evaluation capacities.

- **Advocate a strengthened role of the Executive Board in ensuring a correct implementation of the Evaluation Policy:** There is a need to develop a system that ensures that the evaluation policy is being correctly implemented at the country level. The Executive Board could, as a "reality check", review the overall quality of a (limited) random selection of evaluations commissioned by UNDP's country offices. Further, the Executive Board could oversee an annual visit of a selection of country offices for the purpose of reviewing the conduct of evaluation activities. The possibility of establishing an Evaluation Committee of the Executive Board could be explored as a measure to increase the Board's insight into evaluation activities.

7.3 Issues to be addressed in Sweden's organisation specific strategy for UNDP

It appears appropriate that Sweden's organisation-specific strategy for UNDP places a clear focus on strengthening the role and use of evaluation within the organisation. Hence, it is suggested that the issues addressed above are included in Sweden's organisation-specific strategy for UNDP.

In this respect, it should be noted that Sweden is allocating an important part of its financing to UNDP as earmarked funding. The modality of earmarked funding may contribute to weak governance and management of the organisation, and may hence reduce its relevance and effectiveness. It is therefore critical that earmarked funding from Sweden is appropriately aligned with the organisation's mandate and priorities. Also, it should be consistent with Swedish development goals and the Swedish Strategy's objective of coherent and more results-orientated work and involvement in multilateral development cooperation.

In terms of evaluation, Sweden has an opportunity to allocate a portion of its non-core funding to activities that support the implementation of the UNDP Evaluation Policy at the country level. To the extent that earmarked funding is provided to UNDP, it is proposed that these allocations focus on:

- **The strengthening of in-house evaluation capacities**: This could consist of the financing of training opportunities in evaluation methodology and the development of practical tools for ensuring that project activities are consistent with overarching outcome objectives. The provision of Swedish national experts could be considered in this respect.
- **Further development of the Management Response System:** For ensuring that evaluation findings feed into future activities of UNDP, it appears sensible to focus some ear-marked funding on the further development and institutionalisation of the Management Response System that was recently developed within UNDP.
- **Strengthening of national evaluation capacity building**: Focus could be placed on programmes designed to strengthen national evaluation capacity building. This would, as mentioned above, be consistent with the Paris Declaration, as it strengthens national ownership and, hence, promotes a broader use of evaluations. Related to this is the need to advocate for evaluation: improving the understanding and use of evaluation as a tool for improving performance, and hence increasing development effectiveness. Increasing the demand for evaluation is critical, in order to promote effective national capacity building for evaluation.

References

On meta-evaluation methodology

AEA (2004): *Guiding Principles for Evaluators*, American Evaluation Association

Cars, M. (2006): *A Meta evaluative case study Institute of International Education*, Department of Education, Stockholm University

Forss, K., S. E. Kruse, A. Mwaiselage, A. Nilsdotter, and E. Vedung (2008): *Are Sida evaluations good enough?* An assessment of 34 evaluation reports, A study commissioned by Sida, Stockholm.

Sanders, J.R. (1994): *The Programme Evaluation Standards,* Joint Committee on Standards for Educational Evaluation, Sage, London.

Stufflebeam, D. (2001) *Guiding Principles Checklist for Evaluating Evaluations*, Mimeo

UNFPA (2005): *Strengthening Evaluation for Improved Programming*, UNFPA Evaluation Quality Assessment, UNFPA, New York,

Weiss, C. (1998), *Evaluation*, Prentice Hall, New York

White, H. (2005): *Challenges in evaluating development effectiveness*, IDS Working paper 242, March 2005

Bilateral donors' approaches for assessing multilateral organisations' performance

Cida (2006): *Bilateral Methodologies for Assessing Multilateral Performance: Survey Findings*, Cida

Danida (2008): *Assessing multilateral organisation effectiveness*, Danida

Ministry for Foreign Affairs, Sweden (2007): *Vägledning för bedömning av multilaterala organisationers relevans och effektivitet*, arbetspapper, Government Offices of Sweden, Stockholm

Ministry for Foreign Affairs, Sweden (2007): *Bedömningsmall för Multilaterala Organisationers Relevans och Effektivitet*, budgetprocess inför 2008, Government Offices of Sweden, Stockholm

Ministry for Foreign Affairs, Sweden (2007): *Strategy for Multilateral Development Cooperation*, Government Offices of Sweden, Stockholm

Ministry for Foreign Affairs, Sweden (2007): *UDs faktablad – Biståndet i siffror,* Government Offices of Sweden, Stockholm

MOPAN (2007): *ToR for the MOPAN Working Group: Assessing Agency Performance: Developing a Common Approach*

MOPAN (2007): *Developing a Common Approach for Assessing Multilateral Effectiveness – Phase Two*, Indicator Compendium

NAO, DFID (2005): *Engaging with Multilaterals,* DfID, London

ODI (2005): *Relationships of other donor organisations with multilaterals – Synthesis report*, London

Obser, A. (2007): *Multilateral Organisations Performance Assessment: Opportunities and Limitations for Harmonisation among Development Agencies*

OECD (2007): *Multilateral Organisations' Performance Assessment – Towards a Harmonised Approach*, OECD, Paris

OECD, DAC Evaluation Network (2007): *DAC Evaluation Quality Standards (for test phase application)*, OECD, Paris

OECD (2007): *An Appetiser for the Informal Workshop on Performance and Coherence in Multilateral Development Finance*, OECD, Berlin

OECD DAC Evaluation Network (2007): *New Aid Architecture – Implications for Evaluation*, 6th meeting, OECD, Paris

OECD DAC (2002): *Glossary of Key Terms in Evaluation and Results based Management,* OECD, Paris

RBMG, Results-Based Management Group (2007): *Assessing Multilateral Organisation Effectiveness: A Comparative Analysis of Data Collection Tools*, A discussion paper presented to the MOPAN working group in Stockholm on 11 June 2007

Schater, M. (1999): *Results-Based Management and Multilateral Programming at CIDA: A discussion paper,* IOG

Scott, A. (2004): *Assessing Multilateral Effectiveness*, DfID

Utrikesutskottet, utredningstjänsten (2005): *Utrikesutskottets uppföljning av det multilaterala utvecklings-samarbetet*, 2005/06:RFR6, Riksdagstryckeriet, Stockholm

On UNDP, its evaluation function and the process for becoming a RBM organisation

United Nations (2004): *UN General Assembly Resolution A/RES/59/250,* United Nations, New York

United Nations Country Team (2007): *Common Country Assessment for Nepal* (CCA), United Nations, Kathmandu

United Nations Country Team (2007): *United Nations Development Assistance Framework for Nepal 2008-2010*, 2007. United Nations, Kathmandu

United Nations System in Kenya (2003): United Nations Development Assistance Framework for Kenya 2004 – 2008, 2003. United Nations, Nairobi

United Nations System in Uganda (2004): *Common Country Assessment for Uganda (CCA),* United Nations, Uganda

United Nations System in Uganda (2005): *United Nations Development Assistance Framework UNDAF*, 2006-2010, United Nations, Uganda

UNDP (2002): *Guidelines for Outcome Evaluators*, Monitoring and Evaluation Companion Series #1, p. 9, UNDP Evaluation Office, New York

UNDP (2003): *Second multi-year funding framework*, 2004-2007, UNDP, New York

UNDP (2005): *Draft country programme document for Uganda (2006-2010)*, UNDP, New York

UNDP (2005): *An assessment of UNDP practices in monitoring and evaluation at the country level*, unpublished report, UNDP Evaluation Office, New York

UNDP (2006): *The Evaluation Policy of UNDP*, UNDP, New York

UNDP (2006): *Annual report on evaluation in UNDP in 2006*, UNDP, New York

UNDP (2007): *Regional Workshop on UNDP Evaluation Policy*, 2006 – 07, issues and Recommendations Paper, UNDP Evaluation Office, New York

UNDP (2007): *Draft country programme document for Nepal (2008-2010),* UNDP, New York

UNDP Nepal (2008): Country Programme Action Plan 2008-2010, UNDP, Kathmandu

UNDP (2007): *UNDP Strategic Plan 2008-2011 – Accelerating Global Progress on Human Development*, UNDP, New York

UNDP (2007): *Multi-year funding framework cumulative report on UNDP performance and results for 2004-2006*, UNDP, New York

UNDP (2007): *Evaluation of Results Based Management at UNDP*, UNDP Evaluation Office, New York

UNEG (2007): *The Role of Evaluation in Results-Based Management (RBM),* United Nations Evaluation Group, New York

UNEG (2005): *Norms and Standards for Evaluation in the UN System*, United Nations Evaluation Group, New York

Other

ADB (2008*) Country Operations Business Plan 2008-2010*, Asian Development Bank

ADB (2005): *Nepal: Public Finance Management Assessment*, Strategy and Program Assessment, Asian Development Bank

Bourguignon, F. and M. Sundberg, (2007): *Aid Effectiveness – Opening the Black Box*, American Economic Review, 2007, vol. 97, issue 2, pp. 316-21

Government of Uganda (2005): *Poverty Eradication Action Plan* (2004/5-2007/8), Kampala

Government of Uganda (2005): *The National Integrated Monitoring and Evaluation Strategy Framework – Final Document*, Office of the Prime Minister

Government of Nepal (2003): *The Tenth Plan (Poverty Reduction Strategy Paper)* 2002–2007, Kathmandu

Government of Nepal (2004): *Poverty Monitoring and Analysis System (PMAS) Framework Document*, National Planning Commission, Kathmandu

Government of Nepal (2006): *An Assessment of the Implementation of the 10th plan/PRSP*, Kathmandu

Government of Nepal (2007): *Three Year Interim Plan (2007/08 – 2009/10)*, National Planning Commission, Kathmandu

Government of Kenya (2005) *Poverty Reduction Strategy Paper*, Nairobi

Internet Sources

CGDEV (2007-07-31) *When Will We Ever Learn? Closing the Evaluation Gap*, http://www.cgdev.org/section/initiatives/_active/evalgap

UNDP: *User Guide, Results Management and Accountability*, http://content.undp.org/go/userguide

National Integrated Monitoring and Evaluation Strategy (NIMES), Government of Uganda, http://www.nimes.go.ug/index.php?option=com_frontpage&Itemid=1

Ministry of Finance, Government of Nepal, http://www.ndcm2008.gov.np/

Foreign Aid Coordination Division, Ministry of Finance, Government of Nepal, http://www.mof.gov.np/

International Initiative for Impact Evaluation, http://www.3ieimpact.org

Network of networks for impact evaluation, http://www.worldbank.org/ieg/nonie

Boxes and tables

Appendix I

What approaches for assessing multilateral effectiveness exist today?

Presently, two collaborative approaches including several aid donors in the assessment procedure exist:

- **MOPAN** (Multilateral Organisations Performance Assessment Network).

This is a perception analysis, eliciting the informed judgment of MOPAN member, embassy or country staff (Obser 2007). A MOPAN Working Group was established in 2006 with the objective of working towards a common approach for the assessment of multilateral effectiveness. The Group's priorities are twofold: i) to compare the questions and indicators used in existing methods to assess agency performance and to develop a common core set of questions and indicators that can be used by each MOPAN member, and ii) to work with the multilateral organisations to improve their self-reporting based on a clearer understanding of the information necessary to meet accountability and reporting requirements.

- **COMPAS** (Common Performance Assessment System).

COMPAS is the multilateral development banks' (MDBs) initiative, establishing a common set of indicators in relation to how the MDBs are contributing to development results, and that are suitable for comparative analysis.

Further, a number of bilateral donors have their own corporate systems to track the performance of multilateral operational agencies. Although the overall objective of these models is similar, sources of information and implementation procedures differ considerably.

- **DFID:** The Multilateral Effectiveness Framework (MEFF).

This instrument measures operational effectiveness through a methodology including a checklist and a scorecard to assess eight corporate systems of an organisation, focusing on internal performance, country-level results, and partnerships. Three assessment instruments are used: i) a *checklist* of indicators, expressed as questions, ii) a scorecard rating data in the checklist and iii) a summary report, providing a brief, qualitative review of an organisation's accomplishments (Obser 2007). The MEFF is seen as coming closest to a comprehensive measurement approach, notwithstanding that an important weakness is recognised: the lack of direct coverage of results. Adequate systems and processes are a prerequisite for an effective multilateral organisation but cannot guarantee effectiveness. This means that an organisation complying with the management demands scores well, regardless of the actual results it obtains.

- **Danida:** The Performance Management Framework (PMF).

This approach takes organisational strategies and goals, high-level consultations, and *perception analysis* (MOPAN is one component, but Danida also conducts a separate survey) as a starting point for the analysis.

- **CIDA:** The Multilateral Evaluation Relevance and Assessment system (MERA).

MERA focuses explicitly on three areas in its analysis, against which the organisation will be rated (through the use of *generic indicators*) on a scale from 1 to 5 (1. disastrous, 2. poor, 3. average, 4. very good, 5. outstanding), for: i) relevance, ii) achieved results, and iii) management. Actual results provide the basis for this system, which has proved to be problematic mainly due to attribution problems.

- **Netherlands' Minbuza:** The Multilateral Evaluation Relevance and Assessment system (MMS).

The approach uses a *perception analysis*, providing insights into the general perceptions among embassy staff and their peers of in relation to a given multilateral organisation.

Finally, there is a third approach, addressing perceived deficiencies of both collaborative and bilateral approaches (Obser 2007): the "New approach", initiated by Danida. This approach focuses explicitly on the evaluation function and seeks to answer fundamental questions about the overall quality of an agency's evaluation function. The purpose of the review is to determine whether the evaluation function itself is independent and whether the reports it produces are credible and useful for learning and accountability purposes, as assessed against the UN Evaluation Group (UNEG) norms and standards. To date, UNDP and UNICEF have participated in this peer review process. WFP is the third organisation to be reviewed. The method is currently being further developed through collaboration between DAC and UNEG to establish an internationally-recognised form of peer review of the evaluation function in aid agencies.

In April 2007 the Swedish parliament passed a bill concerning a new comprehensive strategy for Swedish multilateral aid. The strategy is intended to serve as normative guidance for the departments and agencies that interact with the multilateral operational organisations to which Sweden provides funds. It contains an assessment framework in which the multilaterals will be assessed against 1) relevance (i.e. whether the mandate and objectives of the multilaterals coincide with Swedish development cooperation goals), and 2) effectiveness (i.e. whether the multilaterals is contributing to stated goals). The approach is entirely perception-based. It should be stressed that the establishment of this framework is a pilot, constituting an initial step towards a more elaborated model for assessing performance of multilateral organisations.

In summary, MEFF, MERA and COMPAS use specific indicators for performance assessment. PMF and MMS, as well as Sweden's recently established approach, use a measurement process, relying on feedback from representatives in the field following specific questions. MOPAN is similar to the latter approach, and is commonly regarded as "the only game in town", since it involves several donors in a joint initiative on harmonising multilateral performance assessment. Finally, the "New Approach" is distinct again from the other approaches, as it focuses explicitly on the multilateral organisation's evaluation function.

Appendix II

Sample of evaluations

Evaluation sample (25 evaluations)

	Evaluation Title	*Country*	*Region*	*Language*	*Evaluation Type*	*Completion Year*
1	Efficient Administration and Access	Serbia	RBEC	English	Outcome	2004
2	Developing Capacity for negotiations skills and conflict transformation in Zimbabwe	Zimbabwe	RBA	English	Project	2006
3	Revue a mi-parcours du CCF	Côte d'Ivoire	RBA	French	Outcome	2006
4	Democratic Governance Programme	Kyrgyzstan	RBEC	English	Project	2006
5	Final evaluation of Support to Elections outcome	Kyrgyzstan	RBEC	English	Outcome	2006
6	Fortalecimiento de la capacidad de gestión de los gobiernos locales	Ecuador	RBLAC	Spanish	Outcome	2004
7	LA 21-Gender Evaluation	Turkey	RBEC	English	Outcome	2006
8	Enhancing the contribution of international law	Lao	RBAP	English	Project	2004
9	Income generation and sustainable livelihood programme: Outcome Evaluation report	Uganda	RBA	English	Outcome	2005
10	Review of the outcomes of the Governance and Socio-Economic Clusters of Romania's second Country Programme	Romania	RBEC	English	Outcome	2005
11	Amelioration des conditions de vie - Ville de Gabu	Guinea-Bissau	RBA	French	Project	2006
12	Appui a l'Assemblée Nationale	Mali	RBA	French	Project	2004
13	Appui aux Institutions de la Transition (AIT)	Congo DRC	RBA	French	Project	2005
14	Strengthening Economic and Financial Management in the Caribbean Region	Barbados	RBLAC	English	Project	2006
15	Capacity Building Support to the Ministry of Human and Minority Rights	Serbia	RBEC	English	Project	2005
16	Capacity Development for Urban Governance	India	RBAP	English	Project	2005
17	Capacity Development of the National Human Rights Commission of Mongolia	Mongolia	RBAP	English	Project	2005
18	Comprehensive develop for city of Luxor	Egypt	RBAS	English	Project	2004
19	Evaluation of increased public sector efficiency, transparency and accountability	Uganda	RBA	English	Outcome	2005
20	Kenya CO Evaluation Plan - 2006 - 2008	Kenya	RBA	English	Outcome	2006
21	Institute for Training of Public Administration	Albania	RBEC	English	Project	2004
22	National Capacity Building Programme for the Ministry of Foreign Affairs of the Republic of Azerbaijan: The First Phase	Azerbaijan	RBEC	English	Project	2004
23	Strengthening Parliamentary Democracy	Bangladesh	RBAP	English	Project	2006
24	Projet de Renforcement des Capacités des Communes Urbaines(PRCCU)	Burkina Faso	RBA	French	Project	2006
25	Modernizacion del estado pvcia cordoba	Argentina	RBLAC	Spanish	Project	2004

Appendix III

The quality assessment instrument

Title of the evaluation	
Year	
Country/region	
Intervention budget	
Service line within "democratic governance"	– Decentralisation, local governance and urban development – E-governance and access to information – Electoral systems and processes – Justice and Human Rights – Parliamentary development – Policy support for democratic governance – Public administration reform and anti-corruption work – Other
Evaluation type	– project – program – thematic – organisational – outcome – mid-term report – final report
Methods used	– survey – interviews – observations → field visits – document analysis – other:
Instruments used	– qualitative indicators – quantitative indicators – rating scales – benchmarks for assessment – other:
Terms of Reference for the evaluation are attached	
Cost/Size of evaluation	– Financial information, if not: – # weeks spent on evaluation – Categorisation (large, medium, small)
Financing of the intervention	
Evaluated time period	
The evaluation treats:	– Efficiency[36] – Effectiveness[37] – Impact[38] - Relevance[39] – Sustainability[40]

36 Definition according to OECD/DAC Glossary of Key Terms in Evaluation and Results Based Management (2002): A measure of how economically resources/inputs (funds, expertise, time, etc.) are converted to results.

37 Idem (1): The extent to which the development intervention's objectives were achieved, or are expected to be achieved, taking into account their relative importance. Note: Also used as an aggregate measure of (or judgment about) the merit or worth of an activity, i.e. the extent to which an intervention has attained, or is expected to attain, its major relevant objectives efficiently in a sustainable fashion and with a positive institutional development impact.

38 [Idem] Positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended.

39 [Idem] The extent to which the objectives of a development intervention are consistent with beneficiaries' requirements, country needs, global priorities and partners' and donors' policies. Note: Retrospectively, the question of relevance often becomes a question as to whether the objectives of an intervention or its design are still appropriate given changed circumstances.

40 [Idem] The continuation of benefits from a development intervention after major development assistance has been completed. The probability of continued long-term benefits. The resilience to risk of the net benefit flows over time.

	Evaluation criteria	Assessment					
		Very Poor	**Poor**	**Fair**	**Good**	**Excel**	**Missing**
	Overall structure of the report						
1	The report is brief, simple and direct						
2	There is a clear and logical structure to the chapters						
3	There is a clear and adequate executive summary						
4	The annexes are well structured and readable						
5	The report makes use of references						
6	The report is free from grammatical errors						
7	Data collection instruments are carefully described in the evaluation or its annexes						
	Total						
	Purpose, objectives, scope						
8	The purposes of the evaluation…						
9	The objectives of the evaluation…						
10	The scope of the evaluation…						
	Total						
	Evaluator credibility, evaluation independence and ethics						
11	There is an indication of the level of impartiality						
12	Competent evaluators have been engaged						
13	The report discusses evaluation ethics						
14	There is an indication of the level of independence in reporting						
15	There is a discussion on threats to reliability and validity						
16	Possible conflicts of interest are addressed						
17	Stakeholders, that is, persons involved in or affected by the evaluation are identified and have been involved throughout the process.						
	Total						
	Information Sources						
18	The sources of information used						
19	All potential information sources						
20	The evaluation cross-validates and critically assesses						
21	There is a discussion about the limitations of information						
	Total						
	Intervention description						
22	The intervention and its context are described.						
23	The intervention's side effects are identified						
24	The organisational arrangements established						
	Total						
	Evaluation report relevance						
25	The terms of reference are clear and focused						
26	The report is focused on the tasks as defined in TOR						
27	The evaluation question(s) is/are clearly stated						
28	The report is focused on the evaluation questions						
	Total						

	Evaluation criteria	Assessment					
		Very Poor	Poor	Fair	Good	Excel	Missing
	Assessment of evaluation methodology						
29	Methodological choices are described						
30	Methodological choices appear appropriate						
31	The evaluation's limitations are reported						
32	The evaluation report explains the selection of any sample						
	Total						
	Assessment of Evaluation Findings						
33	The intervention's strengths and weaknesses have been assessed..						
34	Methods for assessment of results are specified…						
35	Indicators are used as a basis for results assessment …						
36	Evaluation findings are relevant to the intervention						
37	There is a basis for the relationship between the purpose						
	Total						
	DAC Evaluation Criteria						
38	There is an accurate assessment of efficiency						
39	There is an accurate assessment of effectiveness						
40	There is an accurate assessment of impact						
41	There is an accurate assessment of sustainability						
42	There is an accurate assessment of relevance						
	Total						
	Completeness / Conclusions and recommendations						
43	The evaluation presents conclusions, recommendations and lessons learned separately and with a logical distinction between them.						
44	Conclusions are focused directly on the evaluation questions and each conclusion is supported by findings and/or data						
45	Conclusions are clear and consistent						
46	Recommendations are relevant and targeted to the intended users						
47	Recommendations are practical and may be translated into decisions						
48	Lessons learned are relevant and targeted to intended users						
49	Lessons learned are practical and appears useful						
	Total						
	GRAND TOTAL						

Appendix IV

Guide to the quality assessment instrument

1. Overall Structure of the Report

1. The report is brief, simple and direct
WHAT is the criteria:
The report is **brief**, that is, it treats relevant issues in a concentrate manner. It is **simple**, that is, it follows a logic structure permitting an easy reading. Finally, it is **direct**, that is, it treats relevant issues straight-to-the point, without unnecessary descriptive parts etc.

HOW is the criteria applied:
• 1. *Very poor*. The evaluation does not fulfil any of the three criteria.
• 2. *Poor*. The evaluation does only fulfil some of the three criteria while others are not fulfilled at all.
• 3. *Fair*. Fulfils the three criteria in an adequate manner.
• 4. *Good*. Fulfils the three criteria in a good manner.
• 5. *Excellent*. The evaluation can be considered best practice in this part.
• Missing/not done

2. There is a clear and logical structure to the chapters of the report and to the report as a whole.
WHAT is the criteria:
The report follows a clear and logic structure permitting an easy and efficient reading.

HOW is the criteria applied:
• 1. *Very poor*. There is no apparent structure at all in the report.
• 2. *Poor*. There seems to be little attempts in presenting a report that is clear and logic in its structure.
• 3. *Fair*. Attempts have been made to make a report that is adequately clear and logic in its structure.
• 4. *Good*. The report appears clear with a logical structure, in the chapters and the report as a whole.
• 5. *Excellent*. The evaluation can be considered best practice in this part.
• Missing/not done

3. There is a clear and adequate executive summary
WHAT is the criteria:
The summary provides an overview of the report, highlighting the main conclusions, recommendations and lessons learned. It should be short, concise, readable and well organised – and should “stand alone” (without requiring reference to the rest of the report).
The executive summary should include:
- a brief description of the intervention
- the context of the intervention
- basic description of context and purpose of the evaluation – why this evaluation now?
- Objectives of the evaluation
- Key features and methodology
- Most important findings and conclusions
- Key recommendations

HOW is the criteria applied:
• 1. Very poor. The executive summary does not provide any of the above stated information
• 2. Poor. The executive summary does not provide the whole picture, leaving out essential information, such as key findings, conclusions and/or recommendations
• 3. Fair. The executive summary can stand alone but is missing several elements as stated above.
• 4. Good. The executive summary can stand alone and includes key recommendations, conclusions and findings, but is missing some elements as stated above.
• 5. Excellent. The evaluation can be considered best practice in this part.
• Missing/Not Done

4. The annexes are well structured and readable.
WHAT is the criteria:
All annexes mentioned in the report are attached. These are well structured and readable and provide the information it is said to provide.

HOW is the criteria applied:
• 1. *Very poor*. All mentioned annexes are not attached and those enclosed possess a very low quality.
• 2. *Poor*. All mentioned annexes are attached but are of a very low quality
• 3. *Fair*. All mentioned annexes are attached and are of an adequate quality in terms of structure and information
• 4. *Good*. All mentioned annexes are attached, are well structured and provide the information it is said to provide.
• 5. *Excellent*. The evaluation can be considered best practice in this part.
• Missing/Not Done

5. *The report makes use of references and these are appropriately documented.*
WHAT is the criteria:
Any written material used in the report is well documented and references are made in the text. The bibliography is well structured and includes all necessary information.

HOW is the criteria applied:
• 1. *Very poor*. No references are made in the text, the bibliography is very poorly structured and leaves out important information;
• 2. *Poor*. Some references are made in the text, the bibliography is sufficiently structured but leaves out important information;
• 3. *Fair*. References are sufficiently made in the text; the bibliography is structured and contains adequate information.
• 4. *Good*. References are adequately made in the text; the bibliography is well structured and includes all necessary information.
• 5. *Excellent*. The evaluation can be considered best practice in this part.
• Missing/Not Done

6. The report is free from grammatical and spelling errors
WHAT is the criteria:
The report is written in a clear and direct language, free from grammatical and spelling errors.
This indicator is not possible to divide into 5 different ratings, therefore only the grade 1, 3 or 5 can be given in this section.

HOW is the criteria applied:
• 1. *Very poor*. The language of the report is poor with repeated grammatical and spelling errors
• 2. *Poor*. -
• 3. *Fair*. The language of the report is fair but do contain several grammatical and spelling errors
• 4. *Good*. -
• 5. *Excellent*. The report is written in a clear and direct language and is free from grammatical and spelling errors.
• Missing/Not Done

7. Data collection instruments (protocols, interview questions, list of interviewees etc) are carefully described in the evaluation or its annexes.

WHAT is the criteria:

Data collection instruments (protocols, interview questions, list of interviewees etc) are carefully described in the evaluation or its annexes.
There is a list of persons interviewed, their function and on what specific issue they were interviewed.

HOW is the criteria applied:

• 1. *Very poor*. There seems to be very little attempt to describe any data collection instruments.
• 2. *Poor*. Some attempts have been made for describing data collection instruments but does not provide the whole picture
• 3. *Fair*. Provides an adequate account of data collection instruments.
• 4. *Good*. Provides an good account of data collection instruments. There is a well developed list of persons interviewed, their function and on what specific issues they were interviewed.
• 5. *Excellent*. The evaluation can be considered best practice in this part.
• Missing/Not Done

2. Purpose, objectives, scope of the Evaluation

8. The purposes of the evaluation are described in enough detail, so that they can be identified and assessed

WHAT is the criteria:

The purpose of an evaluation is stated in terms of the evaluation objectives and intended uses of its results. The purpose explains why the intervention is being evaluated.

Corresponding standard in the PES: A3

HOW is the criteria applied:

• 1. *Very poor*. Does not adequately explain why the intervention is being evaluated.
• 2. *Poor*. Does provide some indication why the intervention is being evaluation.
• 3. *Fair*. Provides adequate detail as to why the intervention is being evaluated.
• 4. *Good*. Provides a good account as to why the intervention is being evaluated.
• 5. *Excellent*. The evaluation can be considered best practice in this part.
• Missing/Not Done

9. The objectives of the evaluation are specified

WHAT is the criteria:

The evaluation objectives state what the evaluation seeks to accomplish. Objectives should relate to the purpose and be precisely stated so they guide the evaluator in terms of information needs and data to collect.

Corresponding standard in the PES: A3

HOW is the criteria applied:

• 1. *Very poor*. Does not adequately explain what the evaluation seeks to accomplish
• 2. *Poor*. Does provide some indication of what the evaluation seeks to accomplish
• 3. *Fair*. Provides an adequate account of what the evaluation seeks to accomplish
• 4. *Good*. Provides a good account of what the evaluation seeks to accomplish.
• 5. *Excellent*. The evaluation can be considered best practice in this part.
• Missing/Not Done

10. The scope of the evaluation is clearly defined

WHAT is the criteria:

The evaluation should clearly state its scope in terms of:
a) The issues covered (<u>most important</u>)
b) Funds actually spent*
c) The time period
d) Types of interventions
e) Geographical coverage

HOW is the criteria applied:

• 1. *Very poor*. The evaluation does not adequately define the scope of the evaluation
• 2. *Poor*. Some indication of what the scope of the evaluation is. The issues are covered to a certain extent and some other information regarding the other items (b-e) is provided
• 3. *Fair*. Provides an adequate account of what the scope of the evaluation is.
• 4. *Good*. All items (a-e) are treated. (some indication of funds actually spent)
• 5. *Excellent*. Evaluation meets all of the criterias in a manner that can be considered best practice among the sample of evaluation reports.
• Missing/Not Done

** Rarely mentioned in the evaluations. To receive the mark 4 – good – some kind of information regarding costs or funding has to be presented.*

3. Evaluator credibility, evaluation independence and ethics

11. There is an indication of the level of impartiality in reporting

WHAT is the criteria:

The report is written in a frank and impartial manner and applies different perspectives on issues treated. The reporting procedures should guard against distortion caused by personal feelings and biases of any party to the evaluation, so that the evaluation report fairly reflects the evaluation findings.
Corresponding standard in the PES: P7

HOW is the criteria applied:

• 1. *Very poor*. The evaluation does only give some brief indication on impartiality.
• 2. *Poor*. The evaluation gives some brief indication on impartiality and develops this to a limited extent.
• 3. *Fair*. Provides an adequate account on impartiality but does not develop this to any extent.
• 4. *Good*. Provides an adequate account on impartiality and develops this to a certain extent.
• 5. *Excellent*. The evaluation can be considered best practice in this part.
• Missing/Not Done

12. Competent evaluators have been engaged and the evaluation group possesses a mix of evaluative skills and thematic knowledge.

WHAT is the criteria:

Competent evaluators have been engaged. The evaluation team possesses a mix of:
a) **evaluative skills**, and
b) **thematic knowledge**.
The team is composed of:
c) **international**, as well as **national consultants/ evaluators**,
… and has:
d) A **balanced gender mix**.
Corresponding standard in the PES: U3

HOW is the criteria applied:

• 1. *Very poor*. Incomplete information about the credibility of the evaluators.
• 2. *Poor*. The evaluators possess an unsatisfactory mix of the above stated skills. Some indication of a).
• 3. *Fair*. Provides an adequate mix of skills in terms of a) and b).
• 4. *Good*. Provides an adequate mix of skills in terms of a), b), c), and d)
• 5. *Excellent*. The evaluation can be considered best practice in this part.
• Missing/Not Done

13. The report discusses evaluation ethics
WHAT is the criteria:
The evaluation report should contain a description of the measures and mechanisms put in place to: ensure that the evaluation process was ethical, that stakeholders were protected, and address any ethical dilemmas or issues that emerged. The design should include the following areas:

a) the balance of costs and benefits to participants including potential negative impact,
b) the ethics of who is included and excluded in the evaluation and how this is done,
c) handling of privacy and confidentiality,
d) practices of obtaining informed consent, and
e) feedback to participants

Corresponding standard in the PES: P3

HOW is the criteria applied:
• 1. *Very poor*. Scant attention to ethical considerations (e.g. one or two sentences).
• 2. *Poor*. Some attention to ethical considerations.
• 3. Fair. Provides an adequate description of ethical considerations addressing most of the areas above.
• 4. *Good*. Provides a description of ethical considerations addressing the areas above.
• 6. *Excellent*. The evaluation can be considered best practice in this part.
• Missing/Not Done

14. There is an indication of the level of independence in reporting
WHAT is the criteria:
To what extent are the evaluators independent in their reporting? To what extent is there some information about their independence/dependence? Is there any indication on factors enhancing or threatening their independence?

Corresponding standard in the PES: A3

HOW is the criteria applied:
• 1. *Very poor*. The evaluation does not address any of these questions expressively but some assumptions concerning independence may be made when reading the whole report.
• 2. *Poor*. Addresses some of these questions but in a very brief manner.
• 3. *Fair*. Provides a sufficient indication of the level of independence.
• 4. *Good*. Treats the question of independence in a developed manner.
• 6. *Excellent*. The evaluation can be considered best practice in this part.
• Missing/Not Done

15. There is a discussion on threats to reliability and validity.
WHAT is the criteria:
Reliability refers to the degree of consistency of the information obtained from an information gathering process.
Validity concerns the soundness or trustworthiness of the conclusions that are made from the results of the information gathering process. Validation is the process of compiling evidence that supports the interpretations and uses of the data and information collected using one or more of instruments such as observations, interviews, analysis of documents etc.
Existing and possible threats to reliability and validity should be addressed in the evaluation report.

Corresponding standard in the PES: A5 - A6

HOW is the criteria applied:
• 1. *Very poor*. The evaluation does not explicitly address the issue in the reporting but some notions on the issue may be found.
• 2. *Poor*. Does explicitly address the issue but no further development is made.
• 3. *Fair*. Provides an adequate account on threats to reliability and accountability.
• 4. *Good*. Provides a well developed account on threats to reliability and accountability.
• 5. *Excellent*. The evaluation can be considered best practice in this part.
• Missing/Not Done

16. Possible conflicts of interest are addressed openly and honestly. Measures have been taken to avoid conflict of interests.
WHAT is the criteria:
Conflict of interests exists in an evaluation when the personal or financial interests of an evaluator might either influence an evaluation or be affected by the evaluation.
Conflicts of interest should be dealt with openly and honestly, so that it does not compromise the evaluation processes and results.

Corresponding standard in the PES: P7

HOW is the criteria applied:
• 1. *Very poor*. The evaluation does not explicitly address the issue in the reporting but some notions on the issue may be found;
• 2. *Poor*. The issue has been address, but in a vague and non-straightforward manner;
• 3. *Fair*. The issue is treated in a sufficiently adequate manner;
• 4. *Good*. The issue is treated in a well developed manner;
• 5. *Excellent*. The evaluation can be considered best practice in this part.
• Missing/Not Done

17. Stakeholders, that is, persons involved in or affected by the evaluation are identified and have been involved throughout the process.
WHAT is the criteria:
Who will use the evaluation? Evaluations are mostly used when evaluators work with intended users to define methods and recommendations. Ideally, an evaluation is also intended to be of use to partners and other stakeholders.

Corresponding standard in the PES: P7

HOW is the criteria applied:
• 1. *Very poor*. Hardly any indications on this.
• 2. *Poor*. Users are not clearly defined.
• 3. *Fair*. Provides an adequate account on target groups.
• 4. *Good*. Main users of evaluation are explicitly set out and this is linked to follow-up strategies and recommendations.
• 5. *Excellent*. The evaluation can be considered best practice in this part.
• Missing/Not Done

4. Information sources

18. The sources of information used are described in enough detail, so that the adequacy of the information can be assessed.
WHAT is the criteria:
The evaluators should:

- document, justify, and report their sources of information,
- the criteria and methods used to select them,
- the means used to obtain information from them, and
- any unique and biasing features of the obtained information

… sufficiently to permit others to determine the adequacy of the information for the evaluative questions to be answered. Common errors: Labelling information sources but not describing them.

Corresponding standard in the PES: A4

HOW is the criteria applied:
• 1. *Very poor*. The sources of information are not treated in any of the above stated manners.

• 2. *Poor*. The sources of information are documented but not justified, nor are any of the other elements stated above respected.
• 3. *Fair*. Information sources are fairly documented, justified and reported, however criteria and methods used to select them, the means used to obtain information from them and/or any unique and biasing features of the information are not reported.
• 4. *Good*. Information sources are documented, justified and reported, and 1) criteria and methods used to select them, **OR** 2) the means used to obtain information from them **OR** 3) any unique and biasing features of the information are not reported.
• 5. *Excellent*. Information sources are documented, justified and reported, and criteria and methods used to select them, **AND** the means used to obtain information from them **AND** any unique and biasing features of the information are not reported.
• Missing/Not Done

19. All potential information sources appear to be exhausted.
WHAT is the criteria:
The evaluators should have exhausted all potential sources of information as to maximise the validity and robustness of the findings of the evaluation. A high degree of creativity of the evaluators when scanning the scene for information renders a high mark.

Corresponding standard in the PES: A4

HOW is the criteria applied:
• 1. *Very poor*. Numerous information sources that easily could have fit into the evaluation have been left out for no particular reason
• 2. *Poor*. Most obvious information sources have been exploited, but the approach used by the evaluators was less than creative.
• 3. *Fair*. A sufficiently adequate degree of creativity permeates the reasoning when determining what information sources to be used. Most information sources have been exploited.
• 4. *Good*. A high degree of creativity; a variety of information sources.
• 5. *Excellent*. An excellent degree of creativity; the evaluators have clearly been thinking outside the box and therefore got access to information that enhanced the clarity of the analysis and strengthened the basis upon which the findings of the evaluation rest. The evaluation can be considered best practice in this part.
• Missing/Not Done

20. The evaluation cross-validates and critically assesses the information sources used.
WHAT is the criteria:
The information gathering procedures should be chosen or developed and then implemented in a way to assure that the information obtained is sufficiently reliable for the intended use.

The validity of the data is tested by the use of a variety of methods and sources of information (when feasible).

Corresponding standard in the PES: A6

HOW is the criteria applied:
• 1. *Very poor*. Provides very poor information about validation and assessments of information sources.
• 2. *Poor*. Provides some brief information about validation and assessment of information sources but do not develop this sufficiently.
• 3. *Fair*. Provides adequate information about how information sources have been validated and assessed.
• 4. *Good*. Provides a well developed discussion about validation and assessment of information sources.
• 5. *Excellent*. The evaluation can be considered best practice in this part.
• Missing/Not Done

21. There is a discussion about the limitations of information
WHAT is the criteria:
The evaluation should report potential weaknesses in the data, for example, a single source of information that seemed important but could not be cross-checked or contradictory findings that could not be reconciled.

HOW is the criteria applied:
• 1. *Very poor*. There is very little discussion about the limitations of information used in the evaluation
• 2. *Poor*. There is some discussion about the limitations of information used in the evaluation however weaknesses in the data or contradictory findings are not discussed.
• 3. *Fair*. There is an adequate account on the limitations of information used in the evaluation and weaknesses in the data or contradictory findings are treated to some extent.
• 4. *Good*. There is a well developed account on the limitations of information used in the evaluation and weaknesses in the data or contradictory findings are treated.
• 5. *Excellent*. The evaluation can be considered best practice in this part.
• Missing/Not Done

5. Intervention Description

22. The intervention and its context is described.
WHAT is the criteria:
The context in which the evaluated intervention exists is described in enough detail, e.g. its likely influences on the intervention are identified.

HOW is the criteria applied:
• 1. *Very poor*. The evaluation mentions the context but only very briefly.
• 2. *Poor*. The context is treated but not sufficiently for giving any idea of its likely influences on the intervention.
• 3. *Fair*. Provides an adequate account of the context, permitting an idea of its likely influences on the intervention.
• 4. *Good*. Provides an adequate account of the context as well as its likely influences on the intervention.
• 5. *Excellent*. The evaluation can be considered best practice in this part.
• Missing/Not Done

23. The intervention's side effects are identified
WHAT is the criteria:
Possible side effects that the intervention may have should be identified and reported in the evaluation.

HOW is the criteria applied:
• 1. *Very poor*. The evaluation does only very briefly point out the existence of any side effects.
• 2. *Poor*. The evaluation points out the existence of some side effects but does not develop this further.
• 3. *Fair*. Provides an adequate discussion about the intervention's side effects.
• 4. *Good*. Provides an good discussion about the intervention's side effects
• 5. *Excellent*. The evaluation can be considered best practice in this part.
• Missing/Not Done

24. The organisational arrangements established for implementation of the intervention is described.
WHAT is the criteria:
The evaluation report describes the organisational arrangements established for implementation of the development intervention, including the roles of donors and partners. Measures taken for implementing the intervention are described, such as the establishment of specific contracts, procured consultants, collaborations initiated etc.

HOW is the criteria applied:
• 1. *Very poor*. The evaluation only makes some brief comments about the organisational arrangements.
• 2. *Poor*. Treats the organisational arrangements but not in a sufficient manner.
• 3. *Fair*. Provides an adequate account of organisational arrangements.
• 4. *Good*. Provides an adequate account of organisational arrangements and some information about its consequences etc.
• 5. *Excellent*. The evaluation can be considered best practice in this part.
• Missing/Not Done

6. Evaluation Report Relevance

25. The terms of reference are clear and focused.
WHAT is the criteria:
The ToR should:
- set out the evaluation's intended scope and focus,
- specify evaluation criteria to be used given the evaluation's objective and scope, such as the OECD/DAC criteria,
- explain clearly the outputs and/or products to be produced by the evaluation
- specify the evaluation team and its competencies
- define information sources for data collection,

HOW is the criteria applied:
• 1. *Very poor*. The ToR does not adequately cover any of above stated elements, nor is it clear and focused.
• 2. *Poor*. The ToR covers some of above stated elements.
• 3. *Fair*. The ToR covers all above stated elements but does not develop these sufficiently and/or is not clear and focused.
• 4. *Good*. The ToR covers all above stated elements in an adequate manner and is clear and focused.
• 5. *Excellent*. The evaluation can be considered best practice in this part.
• Missing/Not Done

26. The report is focused on the tasks as defined in the terms of reference.
WHAT is the criteria:
The report should follow the ToR in terms of outputs to be produced by the evaluation, formats for outputs, the methodology, data list and interview list, structure and length of the evaluation report and deadlines.

HOW is the criteria applied:
• 1. *Very poor*. The report does not follow the ToR in any way particular way,
• 2. *Poor*. Does only to some extent follow the ToR
• 3. *Fair*. Refers to the ToR and is adequately focused on the tasks as defined in it
• 4. *Good*. Refers to the ToR and is clearly focused on the task as defined in it
• 5. *Excellent*. The evaluation can be considered best practice in this part.
• Missing/Not Done

27. The evaluation question(s) is/are clearly stated.
WHAT is the criteria:
The ToR need to set out clearly the evaluation's intended scope and focus – that is, the main areas which the evaluation will cover and its main topics of interest.

1. the ToR should either define the main evaluation questions, and/or list objectives
2. Objectives should relate to the purposes and be precisely stated so they guide the evaluator in terms of information needs and data to collect. Evaluation questions should be realistic and achievable
3. The ToR should specify evaluation criteria to be used given the evaluation's objectives and scope, such as the OECD/DAC criteria (relevance, efficiency, effectiveness, sustainability and impact.

HOW is the criteria applied:
• 1. *Very poor*. The evaluation covers only one aspect above.
• 2. *Poor*. Covers only point 1 and 2 partially;
• 3. *Fair*. Covers point 1 and 2 above;
• 4. *Good*. Fully explains the objectives and/or evaluation questions.
• 5. *Excellent.* Fully explains the objectives and/or evaluation questions and presents the key criteria (such as the DAC criteria) providing details on each of the criteria. The evaluation can be considered best practice in this part.
• Missing/Not Done

28. The report is focused on the evaluation questions.
WHAT is the criteria:
The report is structured around and focuses on the evaluation questions. These are treated in a clear and logical manner.

HOW is the criteria applied:
• 1. *Very poor*. The structure of the report has no clear focus on the evaluation questions.
• 2. *Poor*. The structure of the report appears to focus on the evaluation questions to some extent, but these are not treated in a clear and logic manner.
• 3. *Fair*. The report focuses on the evaluation questions and treats these in a sufficiently clear and logical manner;
• 4. *Good*. The report is clearly focused on the evaluation questions and treats these in a clear and logical manner;
• 6. *Excellent*. The evaluation can be considered best practice in this part.
• Missing/Not Done

7. Assessment of evaluation methodology

29. Methodological choices are described.
WHAT is the criteria:
The evaluation report describes and explains the evaluation method and process.

HOW is the criteria applied:
• 1. *Very poor*. The report does not adequately describe nor explain the evaluation method and process.
• 2. *Poor*. The report is only giving a brief indication on evaluation method and process used.
• 3. *Fair*. Provides an adequate account on evaluation method and process .
• 4. *Good*. Provides a well developed account on evaluation method and process .
• 5. *Very Good*. Provides an in-depth explanation on evaluation method and process.
• 6. *Excellent*. The evaluation can be considered best practice in this part.
• Missing/Not Done

30. Methodological choices appear appropriate.
WHAT is the criteria:
The evaluation report describes and explains the evaluation method and process in a manner to permit an independent assessment of its appropriateness by the reader.

HOW is the criteria applied:
• 1. *Very poor*. The report does not adequately describe nor explain the evaluation method and process.
• 2. *Poor*. The report is only giving a brief indication on evaluation method and process used – the reader is not provided enough information permitting an independent assessment of the appropriateness of methods selected.
• 3. *Fair*. Provides an adequate account on evaluation method and process permitting some independent assessment by the reader.
• 4. *Good*. Provides a well developed account on evaluation method and process and discusses the appropriateness of methodological choices.

• 5. *Excellent*. The evaluation can be considered best practice in this part.
• Missing/Not Done

31. The evaluation's limitations are reported.
WHAT is the criteria:
The scope of the evaluation should be clearly stated so that the limitations of the evaluations are clear for stakeholders and intended users.

HOW is the criteria applied:
• 1. *Very poor*. The report does not adequately state the scope of the evaluation and limitations of the evaluation are not clear.
• 2. *Poor*. The report does only briefly state the scope of the evaluation and the limitations of the evaluation are not clear.
• 3. *Fair*. Provides an adequate statement of the scope of the evaluation and the limitations of the same appears clear.
• 4. *Good*. Provides a clear statement of the scope of the evaluation and the limitations are defined.
• 5. *Excellent*. The evaluation can be considered best practice in this part.
• Missing/Not Done

32. The evaluation report explains the selection of any sample. Limitations regarding the representativeness of the evaluation sample are identified.
WHAT is the criteria:
The evaluation report explains the selection of any sample. Changes, during the study, in a sample should be documented and the impact of the changes upon the evaluation results assessed.

Corresponding standard in the PES: A4

HOW is the criteria applied:
• 1. *Very poor*. The evaluation does not adequately describe sources of information, nor is the selection of any sample explained.
• 2. *Poor*. The report does only provide a brief description of information sources and the selection of any sample is not explained.
• 3. *Fair*. Provides an adequate account of information sources and there is a fair explanation of the selection of any sample.
• 4. *Good*. Provides a good account of information sources and there is a good explanation of the selection of any sample.
• 5. *Excellent*. Provides a good account of information sources, the selection of any sample is described and changes, during the study, in the same is documented and the impact of the changes upon the evaluation results assessed. The evaluation can be considered best practice in this part.
• Missing/Not Done

8. Assessment of Evaluation Findings

33. The intervention's strengths and weaknesses have been assessed and reported.
WHAT is the criteria:
The evaluation should be complete and fair in its examination and recording of strengths and weaknesses of the program being evaluated, so that strengths can be built upon and problem areas addressed.

Corresponding standard in the PES: A5

HOW is the criteria applied:
• 1. *Very poor*. Does not adequately assess or report any strengths or weaknesses of the intervention.
• 2. *Poor*. Assesses and reports strengths and weaknesses of the intervention to a very limited extent.
• 3. *Fair*. Provides an adequate account of strengths and weaknesses but do not develop how strengths could be built upon and problem areas addressed.
• 4. *Good*. Provides an adequate account of strengths and weaknesses and develops how strengths could be built upon and problem areas addressed.
• 5. *Excellent*. The evaluation can be considered best practice in this part.
• Missing/Not Done

34. Methods for assessment of results are specified and appear appropriate. (Attribution and contributing/confounding factors should be addressed.)
WHAT is the criteria:
Methods for assessment of results used in the evaluation should be described along with the strengths and weaknesses so that their impact on findings and conclusions can be assessed.

Corresponding standard in the PES: A5

HOW is the criteria applied:
• 1. *Very poor*. Does not adequately describe methods used for assessment of results.
• 2. *Poor*. Describes the methods used for assessment of results but not enough for permitting any assessment of their appropriateness.
• 3. *Fair*. Provides an adequate account of methods used for assessment of results along with their strengths and weaknesses permitting a fair assessment of their appropriateness.
• 4. *Good*. Provides an adequate account of methods used for assessment of results along with their strengths and weaknesses so that their impact on findings and conclusions can be assessed.
• 5. *Excellent*. The evaluation can be considered best practice in this part.
• Missing/Not Done

35. Indicators are used as a basis for results assessment and seem appropriate.
WHAT is the criteria:
In cases when indicators are used as a basis for results assessment their design is sufficiently explained, in order to permit an independent assessment by the reader of their appropriateness.

HOW is the criteria applied:
• 1. *Very poor*. Indicators are being used but these are not explained.
• 2. *Poor*. Indicators are being used; these are explained to some extent but not sufficiently for permitting any independent assessment of their appropriateness.
• 3. *Fair*. Indicators are being used; an adequate account of their design permits a fair assessment of their appropriateness
• 4. *Good*. Indicators are being used; a good account of their design as to permits an assessment of their appropriateness
• 5. *Excellent*. The evaluation can be considered best practice in this part.
• Missing/Not Done

36. Evaluation findings are relevant to the intervention being evaluated and the purpose of the evaluation.
WHAT is the criteria:
The link between evaluation findings and the purpose of the evaluation must be clear and relevant to the intervention being evaluated.

HOW is the criteria applied:
• 1. *Very poor*. Does not adequately clarify the link between evaluation findings and the purpose of the evaluation.
• 2. *Poor*. Does only to some extent clarify the link between evaluation findings and the purpose of the evaluation. The relevance of evaluation findings to the intervention being evaluated is not clear.
• 3. *Fair*. Provides an adequate account of the link between evaluation findings and the purpose of the evaluation. Evaluation findings appear fairly relevant to the intervention being evaluated.

• 4. *Good*. Provides a good account of the link between evaluation findings and the purpose of the evaluation. Evaluation findings appear relevant to the intervention being evaluated.
• 5. *Excellent*. The evaluation can be considered best practice in this part.
• Missing/Not Done

37. There is a basis for the relationship between the purpose of the evaluation, methods used, data collected, and findings clearly presented in the final report.
WHAT is the criteria:
The evaluation makes sense in the way that it is structured. The reader can easily understand how the critical components (purpose, methods, data, and findings) are interlinked and relate to each other. The different components should not feel out of place, but instead contribute to what the evaluation set out to achieve.

HOW is the criteria applied:
• 1. *Very poor*. Does not adequately point out any such relationship.
• 2. *Poor*. Does only to some extent point out such relationship.
• 3. *Fair*. Provides an adequate account of the relationship between the different components of the evaluation.
• 4. *Good*. Provides a good account of the relationship between the components of the evaluation. They are well interlinked and reinforce each other.
• 5. *Excellent*. The evaluation can be considered best practice in this part.
• Missing/Not Done

9. Assessment of how the DAC Evaluation Criteria has been applied

38. There is an accurate assessment of efficiency.
WHAT is the criteria:
Efficiency is a measure of how economically or optimally inputs (financial, human, technical and material resources) are used to produce outputs (*OECD/DAC glossary*).
The evaluation should include the following: assessment of the quality of outputs achieved in relation to the expenditures incurred, and resources used;

a) assessment of timeliness of inputs including personnel, consultants, travel, training, equipment and misc. costs and the timeliness of outputs;
b) whether there was adequate justification for the expenditures incurred and whether the resources were spent as economically as possible, taking into account possible alternatives.

HOW is the criteria applied:
• 1. *Very poor*. Discusses only inputs and does not relate this to achievement of outputs.
• 2. *Poor*. Discusses primarily inputs but with a slight relation to achievement of outputs.
• 3. *Fair*. Assesses quantity, quality and timeliness of inputs and links this to achievement of outputs.
• 4. *Good*. Assesses quantity, quality and timeliness of inputs and links this to achievement of outputs; and analyses whether the resources were spent as economically as possible
• 5. *Excellent*. Assesses quantity, quality and timeliness of inputs and links this to achievement of outputs; and analyses whether the resources were spent as economically as possible, and potential alternative approaches that might have been more efficient. The evaluation can be considered best practice in this part.
• Missing/Not Done

39. There is an accurate assessment of effectiveness.
WHAT is the criteria:
Effectiveness is a measure of the extent to which the intervention's objectives were achieved, or are expected to be achieved, taking into account their relative importance. (OECD/DAC glossary)

The evaluation should include the following:
a) Causality analysis to explain how inputs and activities led to outputs, outcomes and impact. If it is a formative evaluation, it should assess whether inputs and activities are likely to lead to the planned outputs, outcomes and impact;
b) Assessment of coverage (e.g. was the planned areas and target group successfully covered?);
c) Assessment of constraining and facilitating factors and the influence of context on the achievement of results.

HOW is the criteria applied:
• 1. Very poor. Bullet points above not covered at all. Very limited assessment of the interventions effectiveness
• 2. Poor. Bullet points above inadequately covered (e.g.
the evaluation assesses inputs only and/or does not use causality
analysis)
• 3. Fair. Bullet points above covered but only sparsely assessed
• 4. Good. The evaluation provides an assessment of all of the bullet points above.
• 5. Excellent. The evaluation can be considered best practice in this part.
• Missing/Not Done

40. There is an accurate assessment of impact.
WHAT is the criteria:
Impact is the positive and negative long-term effects on identifiable population groups produced by a development intervention, directly or indirectly, intended or unintended. These effects can be economic, socio-cultural, institutional, environmental, technological or other types. (OECD/DAC glossary)

HOW is the criteria applied
• 1. Very poor. Does not attempt to analyze the intervention from the perspective of long-term change, or comment on whether long-term results are likely to be achieved.
• 2. Poor. Gives certain but inadequate consideration to the long-term effects of the intervention.
• 3. Fair. Gives appropriate and visible consideration as to whether long term results have been achieved or are likely to be met. Assessment does not need to be highly rigorous, nor does it need to make a lot of distinction among population groups or the different types of effects listed above.
• 4. Good. Meets satisfactory rating
• 5. Excellent. Meets satisfactory rating, and also applies causality analysis, and provides a better level of analysis about different types of impact. The evaluation can be considered best practice in this part.
• Missing/Not Done

41. There is an accurate assessment of sustainability.
WHAT is the criteria:
Sustainability can be defined as the durability of programme results after the termination of the technical cooperation channelled through the programme. Static sustainability – the continuous flow of the same benefits, set in motion by the completed programme, to the same target groups; dynamic sustainability – the use or adaptation of programme results to a different context or changing environment by the original target groups and/or other groups. (OECD/DAC glossary)

The evaluation should include the following:

a) assessment of the extent to which the programme/project results have had or are likely to have lasting results after programme/project termination and the withdrawal of external resources;
b) assessment of the factors affecting sustainability on the basis of the priority assigned to the programme/project by stakeholders (e.g.. their readiness to continue supporting or carrying out specific activities; replicate the activities in other regions or sectors of the country); or adapting programme/project results in other contexts
c) assessment of the availability of local management, financial and human resources needed to maintain the programme/project results over the long term

HOW is the criteria applied
• 1. Very poor. An evaluation that only states it will be sustainable but does not tell us why, and only partly meets the first two areas above.
• 2. Poor. An evaluation that states that the intervention will be sustainable, but with only with limited analysis/discussion as to why.
• 3. Fair. The likelihood of continued stakeholder support for or adaptation of the programme/project is explained and the evaluation adequately meets the first two areas above.
• 4. Good. The evaluation adequately addresses all three areas above
• 5. Excellent. Includes analysis of local capacity to maintain or adapt programme/project results and adequately addresses all three areas above. The evaluation can be considered best practice in this part.
• Missing/Not Done

42. There is an accurate assessment of relevance.
WHAT is the criteria:
An assessment of programme/project relevance examines the degree to which the outputs, outcomes/purpose and/or goals remain pertinent as originally planned or subsequently modified (OECD/DAC glossary). Analysis should cover:

a) whether the programme/project design is in line with national needs, policies and priorities of programme/project target groups;
b) whether the programme/project is in line with UNDP's policies and priorities
c) synergy between UNDP's intervention and that of other development partners;
d) whether programme/project results are relevant to stakeholders (e.g. have the right kinds of resources, training or information been provided).

HOW is this criteria applied
• 1. Very poor. The analysis meets only the first bullet in a summarized way.
• 2. Poor. The analysis meets the first bullet in a summarized way and addresses bullet d.
• 3. Fair. The evaluation provides an assessment of a c and d.
• 4. Good. The evaluation provides an identifiable assessment of
all bullets.
• 5. Excellent. The evaluation can be considered best practice in this part.
• Missing/Not Done

10. Completeness / Conclusions and recommendations

43. The evaluation presents conclusions, recommendations and lessons learned separately and with a logical distinction between them.
WHAT is the criteria:
The evaluation report should present conclusions, recommendations and lessons learned separately and with a logical distinction between them.

HOW is the criteria applied:
• 1. *Very poor*. Does not adequately separate conclusions, recommendations and lessons learned.
• 2. *Poor*. Does, to some extent, present conclusions, recommendations and lessons learned separately but there is no logic distinction between them.
• 3. *Fair*. Provides an adequate account of conclusions, recommendations and lessons learned. These are presented fairly separately and there seems to be a logic distinction between them.
• 4. *Good*. Provides a good account of conclusions, recommendations and lessons learned. These are presented separately and there seems to be a clear logical distinction between them.
• 5. *Excellent*. The evaluation can be considered best practice in this part.
• Missing/Not Done

44. Conclusions are focused directly on the evaluation questions and each conclusion is supported by findings and/or data.
WHAT is the criteria:
Conclusions are focused directly on the evaluation questions and each conclusion is supported by findings and/or data. The conclusions should add value to the findings of the evaluation.

Corresponding standard in the PES: A10

HOW is the criteria applied:
• 1. *Very poor*. Conclusions do not appear to be adequately focused on evaluation questions and do not appear to be supported by findings and/ or data.
• 2. *Poor*. Conclusions appear to be focused on evaluation questions only to a very limited extent. Conclusions are not clearly supported by findings and/or data.
• 3. *Fair*. Conclusions appear to be adequately focused on evaluation questions and fairly supported by findings and/or data
• 4. *Good*. Conclusions are clearly focused on evaluation questions and strongly supported by findings and/or data
• 5. *Excellent*. The evaluation can be considered best practice in this part.
• Missing/Not Done

45. Conclusions are clear and consistent.
WHAT is the criteria:
The conclusions reached in an evaluation should be clear and consistent. They should be explicitly justified, so that the stakeholders can assess them. Alternative, perhaps even conflicting, conclusions should be reported. Conclusions should flow logically from, and reflect, the report's central findings.

Corresponding standard in the PES: A10 - 11

HOW is the criteria applied:
• 1. *Very poor*. The conclusions reached in the report do not appear to be adequately clear and consistent, nor are they explicitly justified.
• 2. *Poor*. The conclusions reached in the report do not appear to be adequately clear and consistent. They are justified only to a limited extent.
• 3. *Fair*. The conclusions reached appear adequately clear and consistent however their justification is weak.
• 4. *Good*. The conclusions reached appear clear and consistent and well justified.
• 5. *Excellent*. The evaluation can be considered best practice in this part.
• Missing/Not Done

46. Recommendations are relevant and targeted to the intended users.
WHAT is the criteria:
The evaluation report should present recommendations that are relevant and targeted to the intended users of the report. The Recommendations follow logically from the evaluation findings and conclusions and are relevant to the programme/project. They are clearly formulated and they are not too broad or vague.

HOW is the criteria applied:
• 1. *Very poor*. The recommendations given in the report does not appear adequately relevant, nor are they targeted to intended users.
• 2. *Poor*. The recommendations given in the report appear relevant only to a limited extent. They do not appear targeted to intended users.
• 3. *Fair*. The recommendations given in the report appear adequately relevant and appear fairly targeted to intended users.
• 4. *Good*. The recommendations given in the report appear clearly relevant and are targeted to intended users.
• 5. *Excellent*. The evaluation can be considered best practice in this part.
• Missing/Not Done

47. Recommendations are practical and may be translated into decisions.
WHAT is the criteria:
Recommendations are practical and may be translated into decisions.

Corresponding standard in the PES: P7

HOW is the criteria applied:
• 1. *Very poor*. Recommendations do not appear adequately practical. Their translation into decisions seems unclear.
• 2. *Poor*. Recommendations appear practical only to a limited extent. Their translation into decisions seems unclear.
• 3. *Fair*. Recommendations appear adequately practical and their translation into decisions feasible.
• 4. *Good*. The recommendations appear clearly practical and their translation into decisions clearly feasible.
• 5. *Excellent*. The evaluation can be considered best practice in this part.
• Missing/Not Done

48. Lessons learned are relevant and targeted to intended users
WHAT is the criteria:
Lessons learned are relevant and targeted to intended users. The lessons learned make a contribution to general knowledge with implications for future action. Lessons learned are applicable to generic situations and not only to specific circumstances.

HOW is the criteria applied:
• 1. *Very poor*. The lessons learned do not appear adequately relevant, nor are they targeted to intended users.
• 2. *Poor*. The lessons learned appear relevant only to a limited extent. They do not appear targeted to intended users.
• 3. *Fair*. The lessons learned appear adequately relevant and appear fairly targeted to intended users.
• 4. *Good*. The lessons learned appear clearly relevant and are targeted to intended users.
• 5. *Excellent*. The evaluation can be considered best practice in this part.
• Missing/Not Done (ND)

49. Lessons learned are practical and appears useful.
WHAT is the criteria:
Lessons learned are practical and appears useful.

HOW is the criteria applied:
• 1. *Very poor*. Lessons learned do not appear adequately practical. Their usefulness seems unclear.
• 2. *Poor*. Lessons learned appear practical only to limited extent. Their usefulness seems unclear.
• 3. *Fair*. Lessons learned appear adequately relevant and their usefulness fair.
• 4. *Good*. Lessons learned appear clearly relevant and useful.
• 5. *Excellent*. The evaluation can be considered best practice in this part.
• Missing/Not Done

Appendix V

Terms of Reference for CO field visits

SADEV's project on "assessing multilateral effectiveness"

COUNTRY OFFICE FIELD STUDIES

1. Background and purpose of the project

Multilateral development cooperation has been a prominent feature in Swedish development policy for decades. Approximately fifty percent of the Swedish foreign aid budget is channelled through multilateral operational agencies. Out of these fifty percent, agencies within the UN system are receiving about half of the funds. Globally, multilateral operational agencies handle up to a third of total aid volumes. Several of these agencies have a unique role in achieving progress towards the Millennium Development Goals (MDGs).

Research about the effectiveness of multilateral operational agencies has traditionally been sparse. With the ongoing reform of parts of the UN system, the increased focus on Results Based Management (RBM) in aid organisations, the extensive international backing of the MDGs and, last but not least, estimates of an important increase in aid volumes mainly channelled through multilaterals, the issue has, however, recently gained an increased importance in the development community.

Ideally, the assessment and comparison of the effectiveness of a multilateral operational agency should be based on its country level development results. Experience has, however, shown that this is difficult, mainly due to the following reasons: i) the information about the organisation's actual results are often very sparse, ii) the causal link (i.e. the "results chain") between the interventions made and the actual results is often not clear, since final results are influenced on a multitude of different matters (the so-called "attribution problem"), iii) difficulty in aggregating performance measures across projects, sectors and country programs (the so-called "aggregation problem"), iv) the considerable differences between different multilateral organisations making comparison between their results difficult.

Another issue, which further complicates matters, is the fact that many multilaterals provide a wealth of different products or services out of which many are not easily measurable. Activities labelled as either capacity building, advocacy or policy advice are most difficult to quantify and their impact on development objectives is at best noticeable over a longer period of time. [41]

In the absence of reliable information about <u>actual results</u>, the next best alternative has been considered to consist in the assessment of <u>enabling factors</u> necessary for

[41] Scott, A. (2004): *Assessing Multilateral Effectiveness*, DFID, p. 4, Obser, A, (2007) *Multilateral Organisations Performance Assessment: Opportunities and Limitations for Harmonisation among Development Agencies,* DIE, p. 23.

achieving results effectively. The drawback of this approach – not addressing the question of what actually has been achieved – is however commonly recognised.

a. Moving forward…

In the context of the general need for better tools for assessing multilateral operational agencies' effectiveness, several donors have recently pointed out the benefit of an improved use of the multilaterals' self-reporting.[42]

There are several arguments for an increased use of multilateral operational agencies' self-reporting for assessing effectiveness: i) it should encourage consistent and reliable evaluations from these organisations, ii) a well-coordinated process of harmonisation of measurement systems among donors would reduce transaction costs and allow donors to form a common base towards change, iii) harmonisation of measurement practices would be in line with the Paris Declaration iv) it would avoid overburdening organizations with a multitude of donor-specific indicators, and v) it should be an important means of exercising leverage and ensuring greater effectiveness.[43]

Evaluations constitute an important part of multilaterals' self-reporting. If used correctly and consequently, they could have a crucial role in today's results-based culture in establishing credible evidence of the official interventions' impact on the poor.[44] Furthermore, independent and well executed evaluations are crucial for enhancing accountability and hence also organizational learning.

Evaluations as such are (according to OECD DAC vocabulary) *"The systematic and objective assessment of an on-going or completed project, programme or policy, its design, implementation and results. The aim is to determine the relevance and fulfilment of objectives, development efficiency, effectiveness, impact and sustainability. An evaluation should provide information that is credible and useful, enabling the incorporation of lessons learned into the decision–making process of both recipients and donors.*

A prerequisite for an approach using the organisation's evaluations is a **quality assessment** of these evaluations. Such an assessment should focus on determining whether the evaluation at hand covers the aspects characterising an "evaluation" according to above stated definition. That is, that the evaluation - through a "*systematic and objective assessment of an on-going or completed project, programme or policy, its design, implementation and results*" - determines the relevance and fulfilment of objectives, development efficiency, effectiveness, impact, sustainability and, finally, whether the evaluation provides "*information that is credible and useful, enabling the incorporation of lessons learned into the decision–making process of both recipients and donors.*" As previously discussed, development efficiency, effectiveness and impact are tricky to assess in a satisfactory manner. A quality assessment should hence focus on **to what extent** these aspects have been reported, possible gaps in the reporting and **how** reporting on these has been made.

[42] Sweden/Sida (in *Strategy for Multilateral Development Cooperation,* Ministry for Foreign Affairs, March 2007), UK/DfID and Denmark/Danida (in Bilateral Methodologies for Assessing Multilateral Perfromance: Survey Findings, Cida, 2006, p. 12.
[43] Obser, A, (2007), *Multilateral Organisations Performance Assessment: Opportunities and Limitations for Harmonisation among Development Agencies,* DIE, 2007
[44] White, H (2005): *Challenges in evaluating development effectiveness,* IDS Working paper 242, p. 17.

b. SADEVs contribution through the "assessing multilateral effectiveness" project…

The present study is a case-study assessing evaluations carried out by a multilateral operational agency in terms of quality and results reporting. It is focusing on a central sector within one multilateral operational agency's mandate. **UNDP** and its practice area **Democratic Governance** has been selected for the study. Only evaluations carried out at the **project** and **outcome** level will be included into the study. These evaluations are undertaken within the decentralised part of the overall evaluation system. Several important differences between project evaluations and outcome evaluations are evident in their focus, scope and purpose. Outcome evaluations have a broad scope and focus on results to meet a number of overarching goals. Project evaluations have a very limited scope and focus on the processes and inputs of a specific project to better understand that project.[45] Even though tested against the same quality assessment instrument, the assessment will take these differences into consideration and make a clear distinction between these two types of evaluations.

The **objective** of the study is to assess, through a meta-evaluative approach, the quality of a sample of UNDP's Democratic Governance country-level evaluations in terms of:

i) systematic and objective reports,

ii) assessments of the relevance and fulfilment of objectives, development efficiency, effectiveness, impact and sustainability of the evaluated intervention, and

iii) credible and useful information.

The desk review of the sample of evaluation will aim at highlighting strengths and weaknesses in today's reporting as well as finding possible explanations for the variance in quality. Conclusions may be drawn regarding results reporting patterns (output, outcome, impact), the extent to which results have been reported in quantitative/qualitative terms, gaps in the reporting, etc. Persisting deficiencies in the reporting and possible solutions to these will be discussed, based on the findings of the study.

A number of country offices of the agency will constitute the focus for a more **in-depth and comprehensive study**. This part will be more **process focused** and will include a larger scope of information sources than the previous desk review. UNDP's country offices in **Kenya**, **Uganda** and **Nepal** are currently considered for inclusion.

2. Country Office field studies

a. Objective

The Country Office field studies will be more **process focused** than the desk review. The objective of these, more in-depth, studies is to identify useful and concrete examples of good practice as well as key weaknesses in evaluation practice. This should result in conclusions on factors potentially affecting the quality of, and results

[45] UNDP (2002): *Guidelines for Outcome Evaluators*, Monitoring and Evaluation Companion Series #1, p. 9, New York, UNDP Evaluation Office

reporting in, country office evaluations and may permit some triangulation of the findings from the desk review. In general, the aim of the field studies is to provide insights about evaluation capacity at UNDP COs. In this context it is important to consider the factors that drive the demand for COs to produce good quality evaluations.

b. Scope

The Country Office field studies will include a larger scope of information than the desk review. First of all, **all evaluations** (i.e. not, as will be the case for the desk review, only evaluations within the democratic governance practice area) submitted to the ERC concerning these countries will be included into the study. Additional documentation, such as evaluations existing at the country offices archives but not available in the ERC should, to the extent possible, be included.

A number of interviews will be conducted with evaluators, commissioners and different stakeholders to the evaluations included into the study (see below).

c. Methodology

The Country Office field studies will include i) a **desk review** of relevant documentation, including a test of available evaluations against the quality assessment instrument used for the desk review, and ii) **extended interviews** with evaluators, commissioners, programme partners, and different stakeholders to the evaluations included into the study.

i. Country Office staff

Interviews will be conducted with CO staff responsible for i) evaluations and ii) UNDP's projects and programmes.

ii. Other stakeholders

Interview should, to the extent possible, be conducted with other interest groups and stakeholders to the evaluations, such as Government officials, CSO representatives, independent consultants etc.

d. Themes/ Issues to be addressed

Themes that will be addressed through semi-structured interviews include:

- General evaluation activity / factors that drive demand for evaluations in COs: who decides what to evaluate? When does an evaluation get commissioned? What kind of evaluations is most common?
- Evaluation capacity – guidelines/ interaction/ directions from EO. Different Types of Evaluations; awareness of Outcome Evaluation methodology. Methodological advances amongst CO staff within evaluation?
- Evaluator competencies – recruitment of consultants
- Evaluation prerequisites: the linkage between evaluations and the design of programs

- Evaluation partnerships: efforts for strengthened ownership, alignment, harmonisation, managing for results and mutual accountability, in accordance with the Paris Declaration.
- Results reporting
- Evaluation use / Evaluation follow-up: management response

e. Structure of the country office field studies reports

A common approach to all Country Office field studies will be ensured following a standardised structure, briefly outlined below:

- Country context
- UNDP involvement in the country
- UNDP country programme (content and time frames)
- UNDP Evaluation plans (- how do UNDP evaluations and evaluation process coincide with the country's priorities?)
- Evaluation: Process
- Evaluation: Product
- Conclusions

f. Evaluation ethics

The Country Office field studies will be undertaken with integrity and honesty; they will be designed and conducted with respect to the rights and welfare of human subjects involved.

g. Amount of time for a country office field study

The conduct of a Country Office field study is estimated to take **3 – 4 days**. The total amount of time spent at a country office is however obviously depending on the availability of relevant interviewees as well as the country office's general capacity to assist the SADEV team.

h. Time schedule / time frame

The finalisation of the project is set to the end of **February / early March 2008**.

Two missions, including visits to two different Country Offices each, are planned for. Ideally, the first mission would take place at the end of November / early December 2007, and the following mission during January 2008. The SADEV team is well aware that the staff of UNDP has tight schedules. In case that the proposed dates are not feasible, SADEV is open to discuss other dates.

Appendix VI

Kenya Mission Schedule

Time	Meeting	Participants	Venue
MONDAY 17 December 2007			
Time	*Meeting*	*Participants*	*Venue*
9.00 – 10.00	UNDP Country Programme	Nardos Bekele-Thomas DRR(P), Sheila Ngatia ARRai, Rose Muchiri M&E	UNDP
10.00 – 10.30	UNDP Programme M&E	Rose Muchiri	UNDP
10.30 – 11.30	Human Rights Programme and Elections Programme	Sheila, Maina , Enid, Elisabet, Janneke	UNDP
12.00 – 1.00	Kenya Private Sector Alliance		KEPSA
1.00 – 2.00	Lunch		
2.00 – 2.45	Federation of Kenya Employers		FKE
3.00 – 4.00	MDG Implementation Unit, Ministry of Planning and National Development and M.E. Director	Gideon Mailu, Head of MDG Implementation Unit Mr. Mwando, MED	Treasury Building
4.00 – 5. 00	External Resources Department, Ministry of Finance	Bernard Masiga, UN Desk Officer	Treasury Building
TUESDAY 18 December 2007			
Time	*Meeting*	*Participants*	*Venue*
9.00 – 10.30	Public Service Reform Programme	Mary Njoroge, Head of PMU, Mavis Nathoo, Head of TAU	PSRDS
11.00 – 12.00	CPAP MTR Evaluation	Joyce Deloge and John Njoroge, Evaluators	UNDP
12.00- 1.00	Human Rights and Civic Education Evaluation	Mr. Karuti Kanyinga, Evaluator	UNDP
1.00 – 1.30	Lunch		
2.00 – 3.00	National Civic Education Programme (NCEP)	Mr. Zein Abubaker, NCEP Coordinator	NCEP offices
3.30 – 5.00	Sida	Sida	Sida
WEDNESDAY 19 December 2007			
Time	*Meeting*	*Participants*	*Venue*
9.00 – 9.45	Action Aid	Programme & M&E Officers	Action Aid
10.00 – 10.45	Clarion	Programme &M&E Officers	Clarion
11.00- 11.45	Legal Resources Foundation	Programme & M&E Officers	LRF
12.00 – 12.45	Kenya Human Rights Commission	Programme & M&E Officers	KHRC
1.00-1.30	Lunch		
2.00 – 2.45	National Council for Persons with Disabilities		NCPWD
3.00- 3.45	United Disabled Empowerment Kenya		UDEK
4.15 – 4.45	Debriefing with UNDP	UNDP	UNDP

Uganda Mission Schedule

Agenda for the SADEV Mission to UNDP Uganda

Date	Time	Activity	Venue	Focal Point	Status
Monday 14th January	9:30 AM	Meeting the Resident Representative	UNDP	Jenifer Rusiita	Confirmed
	10:00 AM	Meeting Governance Unit staff	UNDP	John Mpande	Confirmed
	2:30 PM	Meeting with PDCO	PDCO Offices	Okumu Dison	Confirmed
	4:00 PM	Meeting with PPDA Director - Procurement Authority	UNDP	John Mpande	Pending
Tuesday 15th January	9:00 AM	Meeting evaluation consultants of DG & CPR	UNDP	John Mpande	
		James Katorobbo		John Mpande	confirmed
		Frank Muhereza		John Mpande	confirmed
	11:00 AM	Meeting with PDM Coordinator (MoLG)	Workers House	Mr. Kayuza Joseph	confirmed
	12:00 PM	Meeting with AMICAALL Coordinator	Workers House	Dr. Mugisa	confirmed
	2:00 PM	Meeting with the IGG	IPS Building	Mrs. Kaberuka Jane	confirmed
	4:00 PM	Meeting with Ms. Rose Nalwadda	UAC	Ms. Nalwadda	confirmed
Wednesday 16th January	9:00 AM	Meeting with Commissioner ALD	ALD	John Mpande	
	11:00 AM	UNDP Programme meeting	UNDP	John Mpande	

Nepal Mission Schedule

Time		Venue	Remarks
Sunday, 20 January			
1145 hrs	Arrive by Jet Airways - 9W 262 and check-in at Summit Hotel	Hotel vehicle will pick up from airport.	
Monday, 21 January – Meetings with UNDP			
1300 hrs	Meeting with Ms. Heather Bryant, Monitoring and Evaluation and Knowledge Management Officer	SCN's office	Confirmed
1730 hrs	Meeting with Ms. Anne-Isabelle Degryse-Blateau, Country Director; Mr. Ghulam Isaczai, Deputy Resident Representative; Mr. Rahama Mohammed, Deputy Resident Representative/ Operations	CD's office	Confirmed
Tuesday 22 January – Meetings with Government and other Counterparts			
0930 hrs	Meeting with Dr. Jagadish Chandra Pokharel, Vice-Chairman, National Planning Commission (*Note: Dr. Pokharel was also a member of the Outcome Evaluation team for "Pro-Poor Policies")*	NPC	Confirmed
1200 hrs	Meeting with Mr. Ganga Dutta Awasthi, Acting Secretary, Mr. Bishnu Nath Sharma, Joint Secretary, Mr. Som Lal Subedi, Joint Secretary, Ministry of Local Development (*Evaluations of DLGSP & DFDP; Outcome Evaluation)*	MLD	Confirmed
1400 hrs	Meeting with Ms. Padma Mathema, Secretary, National Human Rights Commission and National Project Director for CDNHRC Mr. Keith Leslie, Sr. Human Rights Adviser, Mr. Tek Tamrakar, National Programme Manager, CDNHRC *(Evaluation of CDNHRC)*	NHRC	Confirmed
1500 hrs	Meeting with Team of Commissioners, National Human Rights Commission *(Evaluation of CDNHRC)*	NHRC	Confirmed
Wednesday, 23 January – Valley strike			
0830 hrs	Meeting with Ms. Heather Bryant, Monitoring and Evaluation and Knowledge Management Officer	UNDP	Confirmed
Thursday, 24 January – Meetings with members of evaluation teams and other stakeholders			
0930 hrs	Meeting with Ms. Kanta Singh, National Consultant *(Member of DLGSP Evaluation Team)*	UNDP SRP's office	Confirmed
1400 hrs	Meeting with Mukunda Raj Prakash, Under Secretary, Ministry of Local Development *(Government Member of DFDP Evaluation Team)*	MLD	Confirmed
1530 hrs	Meeting with Mr. Krishna Prasad Sapkota, Chairman and Mr. Hem Raj Lamichane, Executive General Secretary, Association of District Development Committees of Nepal (ADDCN) *(Key partner in local governance)*	ADDCN	Confirmed
1700 hrs	Meeting with Mr. Sharad Neupane, Assistant Resident Representative	SCN's office	Confirmed
Friday, 25 January – Additional Meetings			
0900 hrs	Meeting with Mr. Hemang Sharma, National Consultant *(Member of CDNHRC Evaluation Team)*	UNDP Conference Room	Confirmed
1100 hrs	Meeting with Mr. Subarna Lal Shrestha, Joint Secretary and Ms. Mandira Poudel, Programme Director, National Planning Commission	NPC	Confirmed

Time		Venue	Remarks
1400 hrs	Meeting with Mr. Ram Krishna Timalsena, Registrar, Supreme Court	Supreme Court	Confirmed
1630 hrs	Debriefing with Ms. Anne-Isabelle Degryse-Blateau, Country Director and Mr. Ghulam Isaczai, Deputy Resident Representative	CD's office	Confirmed
Wednesday, 30 January			
1100 hrs	Meeting with Ms. Neeta Thapa, National Consultant (Member of DFDP and other Evaluation Teams)	UNDP	Confirmed
1715 hrs	Depart by Etihad Airways – EY 291		Hotel vehicle will drop at airport

Appendix VII

Interview guide

The interviews were framed around a set of questions, depending on what category of stakeholder that was being interviewed. The different stakeholders were: 1) UNDP Country Office Staff, 2) Government Officials, 3) Independent Evaluation Consultants, and 4) other stakeholders (such as representatives for various NGOs)

UNDP Country Office Staff

1. Evaluation capacity

- Do you have extensive evaluation capacities at the CO itself, and what mandate of initiating/producing your own evaluations do you have?
- Is there any M&E officer at place?
- The availability of competent evaluators locally - Is there any external evaluation network/association at regional or national level?

2. Evaluation prerequisites

- Who defines and develops measurable objectives and appropriate indicators?
- Is there a clear link between outcome evaluations and the way DG programmes generally are designed?
- Are there any efforts made for jointly agreed indicators, terminology and reporting formats, nationally and with other UN agencies/donors?
- What is your opinion on assessing progress through outcomes?

3. Results reporting

- What are the practical implications of moving from being a results-orientated organisation to an organisation using indicators of outcomes?
- Has results-based planning and systems development been internalised within country programming?
- What are the main challenges in handling outcome evaluations?

4. Evaluation use

- Changes at CO level since the launch of the tracking system in 2006 (Management response system)
- How is the knowledge generated by evaluations fed back into the organisation? Locally and on a central level?

5. Evaluation partnerships

- How do UNDP evaluations and evaluation relate to the country's priorities?
- Correlation between evaluations and evaluation processes with the country's own priorities
- What are the challenges in your agency/country when making joint evaluations?
- Degree of stakeholders' (target groups) participation in evaluation
- Interaction between UNDP and the State in the field of monitoring and evaluation at the country level; engagement with local and national capacities.

Government Officials

1. General evaluation activity

- National ownership of evaluation… how does this work in practice?
- What is your experience of working jointly with UNDP and/or other UN Agencies?

2. Evaluation capacity

- UNDP's or other UN Agencies' role in assisting the Government in the development of evaluation capacity
- Does government mainly use internal or external evaluators?
- Is there any external evaluation network/association at regional or national level?

3. Evaluation prerequisites

- Are there any efforts made for jointly agreed indicators, terminology and reporting formats, nationally and with UN agencies/donors?
- How do donors' (UNDP) evaluations and evaluation process relate to your country's priorities?

4. Results reporting and evaluation use

- Has results-based planning and systems development been internalised within country programming?
- How do you generally make use of findings in evaluations? To whom are they reported?
- Do external evaluations from agencies like UNDP feed into departmental policy making?

Evaluation Consultants

1. General evaluation activity

- Around how many evaluations do you do per year?
- Which are your most important clients?
- Is there any external evaluation network/association at regional or national level?

2. Evaluation capacity

- Formal training, background, experience
- What kind of guidance did your scopes of work or contract give you about the evaluation methods to use? – Are they generally sufficient and clear?
- Describe your relationship with UNDP during the evaluations, e.g did you have adequate independence? Were they supportive? Where they defensive?

3. Evaluation prerequisites

- Is there usually clarity in programme objectives and in sufficiently defined performance indicators?
- Are there any efforts made for jointly agreed indicators, terminology and reporting formats, nationally and with UN agencies/donors?
- Are project reports designed or adapted to create evidence geared towards measuring development results?

4. Evaluation partnerships

- Do your evaluations usually include participation of stakeholders (target groups)?

- Is it common with Inter-agency collaborations in the M&E area, and what would be your role in that partnership?

5. **Results reporting and management response**
- What is generally the format or presentation of knowledge generated by evaluation?
- Do you feel that the findings in your evaluations are used to influence decision making?

Other stakeholders:

1. **General evaluation activity**
- Do you generally evaluate your programs and projects? At what level and what kind of evaluation do you prefer?

2. **Evaluation capacity**
- Are your evaluations generally made in house or by external consultants?
- The availability of competent evaluators locally - Is there any external evaluation network/association at regional or national level?
- Do you cooperate with other NGOs or with UN agencies in using evaluators?

3. **Evaluation prerequisites**
- Are there any efforts made for jointly agreed indicators, terminology and reporting formats, nationally and with other UN agencies/donors?
- Who defines and develops measurable objectives and appropriate indicators?
- Is there a clear link between outcome evaluations and the way DG programmes generally are designed?

4. **Evaluation partnerships**
- How do UNDP evaluations and evaluation process coincide with your organisation's priorities?
- Are you frequently partners with UNDP or other donors in evaluations?
- What characterises your partnership or how you cooperate with international donors within evaluations?

5. **Results reporting**
- How are your project reports and evaluations fed into a national or wider data set?
- Is there any form of partnership with government or UN agencies in gathering and processing results?

6. **Evaluation use**
- What is the format or presentation of knowledge generated by evaluation?
- Main use of evaluations; intended users, reasons for conducting evaluations
- Do other actor's results and evaluations feed into your decision-making process?

Appendix VIII

Development context in Country Offices visited

1. Kenya

The field visit to Kenya was conducted in December 2007, just one week before the general elections that ultimately led to serious political turbulence and violence in the country. Clearly this political turmoil and associated violence would affect all actors that work through public institutions, including UNDP.

Foreign Aid in Kenya – brief about Kenya as receiver of foreign aid

Donor assistance is less important in Kenya than in many other African countries, although it remains important in some sectors. The mean value for ODA as a share GDP in Kenya over the period 2000-2006 was 2.5 per cent.[46] However, for some activities, such as HIV/Aids, education and water supply, development partners are still funding a significant proportion of the resources.[47]

Aid climate – guiding strategic documents

Much remains to be done to reach the targets of the Paris Declaration. Many donors in Kenya are still not using government procurement and financial management systems, field missions and analytical work are conducted independently of other development partners, and independent project implementation units are continuously relied upon to manage interventions.[48]

Nonetheless, donor coordination has improved since 2003. Since 2004, the government and donors have been engaging in regular meetings within the Kenya Coordination Group, chaired by the MoF, to discuss matters of mutual concern. Development partners meet among themselves each month in the Development Coordination Group, chaired by the World Bank. Furthermore, the Harmonisation, Alignment and Coordination Group, which includes the MoF, actively promoted the aid effectiveness agenda. In 2007, all 17 of its members,[49] covering some 90 per cent of total official development assistance to Kenya, joined together to formulate the Kenya Joint Assistance Strategy (KJAS). This strategy is aligned to the government's strategy Kenya Vision 2030 (see below).

In the period 2008-2009, Kenya will be entering into a new phase of programmatic documents and strategies: UNDAF 2004–2008 will be replaced by UNDAF 2009-2013 (3rd version) which will found the basis for the new CPAP (2009-2013). In drafting this UNDAF, regard was taken of the government's newly established strategy for development, the Kenya Vision 2030, which will enter into force in 2008. The Kenya Vision 2030 is to be implemented in successive five-year Medium Term Plans, with the first such plan covering the period 2008-2012. A detailed 5-year

[46] Numbers from DAC Foreign Aid Statistics and World Economic Outlook, IMF
[47] Kenya Joint Assistance Strategy, p 17.
[48] [Idem]
[49] Members to the KJAS are: Canada, Denmark, the European Commission, Finland, France, Germany, Italy, Japan, the Netherlands, Norway, Spain, Sweden, UK, the AfDB, the UN, and the World Bank.

development plan (2008-2012) has recently been prepared under the coordination of the Ministry of Planning and National Development.

Evaluation Climate[50]

The government has proved committed to the results-based approach to development and has made efforts to collect and report on information that will allow stakeholders to assess the effectiveness of the government's strategy in achieving its development objectives and the MDGs. An integrated sector-wide monitoring and evaluation (M&E) system has been developed as part of the government's programme to improve the allocation and use of resources, in order to enhance governance, transparency and accountability. This system will serve as an essential tool for increasing the efficiency and effectiveness of public policies in achieving the Economic Recovery Strategy (ERS) objectives, including the Millennium Development Goals (MDGs) and the Kenya Vision 2030.

A Monitoring and Evaluation Department within the Ministry of Planning and National Development was established in 2004. It is responsible for collating, coordinating, and disseminating information. The government has also created a three-tier (MDG, African Peer Review Mechanism (APRM), ERS) framework for monitoring and evaluating the implementation of the ERS and its effectiveness in stimulating growth and reducing poverty and inequality.

The government has continued to revitalise its evaluation and monitoring capacity. A National Monitoring and Evaluation Steering Committee has been established, comprising government stakeholders from the Ministries of Planning and National Development (including from the CBS), the Ministry of Finance, development partners, and nongovernmental organisations. The Cabinet Office has strengthened its capacity to track the implementation of cabinet decisions, presidential pronouncements, and government policies.

In collaboration with its development partners, the government continues to institutionalise and strengthen its M&E system by enhancing capacity for generating statistics through the multi-donor National Statistical System Project, articulating a nationwide framework, and establishing links with planning tools such as the budgeting process, medium-term expenditure framework (MTEF), and the public expenditure reviews.

In 2004, Kenya launched the MDG-based planning and policy formulation process to mainstream the MDGs within the national policy, budgeting, and monitoring processes of the government.

The government is making progress in strengthening its capacity for M&E, however much remains to be done. The capacity of institutions needs to be strengthened, together with a framework for coordinating activities of the many actors involved in M&E in the country.

---

[50] This section is based on information provided at http://siteresources.worldbank.org/INTKENYA/Resources/me_ers_donor.pdf as well as on interviews conducted in Kenya, December 2007.

2. Uganda

Foreign Aid in Uganda – brief about Uganda as receiver of foreign aid

Uganda is one of the poorest countries in the world, and is heavily dependent on foreign aid to finance public sector programmes. The mean value for ODA as a share GDP in Uganda over the period 2000-2006 was 8.9 per cent.[51] Foreign aid funds some 50 per cent of public expenditure. Nonetheless, over the last 15 years the country has experienced considerable GDP growth, which has averaged over 6 per cent annually.[52]

Aid climate – guiding strategic documents

The principal strategic document guiding public actions is the Poverty Eradication Action Plan (PEAP). The PEAP is fully endorsed by GoU, and is the national development framework to which all external donors (bilateral and multilateral) must conform. It articulates the policies adopted by the GoU for making progress towards various developmental indicators. The PEAP was first introduced in 1997 and has been revised every few years to reflect implementation progress, and to adjust medium-term policies to the achievement of long-term objectives. In line with the Paris Declaration, monitoring and evaluation of development programmes and projects should relate to the performance indicators as expressed in the Policy and Results Matrix of the PEAP.

Evaluation Climate

The Government plays a strong role in coordinating external assistance. The Aid Liaison Department (ALD), which is situated within the Ministry of Finance, Planning and Economic Development (MFPED), is responsible for aid coordination.

Effective use of evaluative evidence within public sector management has proven problematic in Uganda. To date, managers in the public sector are more concerned with bureaucratic accountability for expenditures than in being accountable for development results[53]. Major challenges to effective M&E include weak coordination arrangements, parallel M&E efforts, poor public management culture, gaps in information, and under-utilised information.[54] Nonetheless, there is evidence of public resolve to improve existing M&E systems. The Office of the Prime Minister is responsible for coordination, and the monitoring and evaluation, of national and sectoral policies and programmes. This is done to ensure coherence and internal consistency with the overarching development strategy.[55] Situated within Office of the Prime Minister, the National Integrated Monitoring and Evaluation Framework (NIMES), is a coordinating M&E unit, but is not an operational M&E system in its own right. Hence, it is a framework for coordination, with the ultimate goals of

[51] Numbers from DAC Foreign Aid Statistics and World Economic Outlook, IMF
[52] Terms of Reference for Mid-Term Independent Evaluation of Uganda's Poverty Eradication Action Plan, GoU, Kampala
[53] The National Integrated Monitoring and Evaluation Strategy Framework – Final Document (2006), GoU, Office of the Prime Minister
[54] PEAP 2004-2007
[55] NIMES website (080403)
http://www.nimes.go.ug/index.php?option=com_content&task=category&sectionid=13&id=28&Itemid=37

reducing duplication of M&E efforts and increasing the quality and utility of developmental data.[56]

NIMES aims to establish a single M&E system for both government and donors alike. This system will draw from existing reporting mechanisms for sector-specific support, the PEAP and the budget process. The Policy and Results Matrix of the PEAP is used for performance assessment, and will serve as the benchmark for assessing annual progress and feed into government progress status reports.

3. Nepal

Foreign Aid in Nepal – brief about Nepal as receiver of foreign aid

Foreign aid plays a significant role in Nepal's development. The mean value for ODA as a share of GDP in Nepal over the period 2000-2006 was 4.3 per cent.[57] [58] In the financial year 2005-06, total foreign assistance made up almost 60 per cent of the total allocation for development programmes.[59] Over the past four years the economic growth rate averaged 3.3 per cent of GDP, which was slightly lower than the PRSP predicted.[60]

Aid climate – guiding strategic documents

2007 marked the end of the latest PRSP – the tenth government plan – which covered the preceding six year period. As the country is in a state of transition, the government decided that it would produce an interim 3-year plan, from mid-2007 to mid-2010, to be followed by the Eleventh Plan in 2010, which will have a five year planning horizon, the same as the tenth plan. The interim plan articulates the GoN's intention to address development challenges during this transitional period. The government of Nepal faces enormous challenges in advancing the implementation of development Programmes, particularly in the context of the country's recent upheavals.[61] The PRSP review of 2005-06 concluded: "… in order to implement the Paris Declaration, a systematic programme is needed to restrict aid effectiveness commitments, developing national action plans, setting baselines and introducing more effective monitoring systems."[62]

Evaluative Climate

The Ministry of Finance is the lead institution responsible for coordinating external assistance, in the form of foreign aid through the Foreign Aid Coordination Division (FACD). FACD does not evaluate itself or the public sector programmes that foreign aid is funding.[63] Rather, the division acts as focal point for the line ministries for foreign-assisted projects. Thus, FACD is involved in the life-cycle of all development projects: from the preparation stage, to negotiation, approval and actual implementa-

[56] The basic idea is that NIMES will integrate PEAP M&E, sector ministries' information systems, and financial management systems under a single framework.
[57] Kenya Joint Assistance Strategy, p 17.
[58] Numbers from DAC Foreign Aid Statistics and World Economic Outlook, IMF
[59] Government of Nepal (2006): An Assessment of the Implementation of the 10th plan/PRSP,
[60] [Idem]
[61] GoN (2007) Three Year Interim Plan (2007/08 – 2009/10), Kathmandu, National Planning Commission
[62] Government of Nepal (2006), supra note 55
[63] Ministry of Finance, GoN, http://www.ndcm2008.gov.np/

tion.[64] The National Planning Commission (NPC) provides programming direction and is the coordinating governmental body for M&E activities. The NPC provide follow-up and analysis of progress that has been made in relation to the indicators set out in the PRSP. A separate poverty-monitoring unit is situated within the Monitoring and Evaluation Division of the NPC.[65] The NPC receives quarterly performance reports from line ministries. Information is channelled through sector ministries, which then submit performance data to NPC on a quarterly basis. Over the past four years the NPC has championed a process of institutionalising a poverty monitoring and analysis system (PMAS) framework. The PMAS works at the national system level, but a district level monitoring system is currently in planning stage.[66] The PMAS was essentially set up to support the implementation of the PRSP.[67] A number of steps have been taken to adopt an RBM approach within government to manage the public sector.[68]

Although there are indications of progress in M&E in Nepal, considerable obstacles must be overcome for M&E to work properly and to become an active part of public sector management. The limited institutional capacity of the government affects all facets of programming, including issues related to M&E.[69]

Current problems associated with ineffective M&E include: 1) low capabilities in interpreting development data and utilising M&E findings for decision making, 2) weak coordination of M&E and inadequate harmonisation of donor activities, and 3) ineffective human resource management.[70]

[64] FACD website 080403: http://www.mof.gov.np/

[65] ADB (2005): Nepal: Public Finance Management Assessment, Strategy and Program Assessment, Asian Development Bank

[66] Government of Nepal (2006), supra note 55

[67] Government of Nepal (2004): Poverty Monitoring and Analysis System (PMAS) Framework Document, Kathmandu, National Planning Commission

[68] ADB (2008) Country Operations Business Plan 2008-2010, Asian Development Bank

[69] ADB (2005), supra note 61

[70] [Idem]

Appendix IX

List of interviewees

Kenya Mission 17-20 December 2007

UNDP Country Office	
Ms. Nardos Bekele-Thomas	Deputy Resident Representative
Ms. Sheila Ngatia	Assistant Resident Representative, Empowerment Unit
Ms. Rose Muchiri	Monitoring and Evaluation Specialist
UN Resident Coordinator	
Ms. Inderpal Dhiman	Head of UN Resident Coordinator Secretariat
Ministry of Planning and National Development, MDG Implementation Unit, GoK	
Mr. Gideon Mailu	Head of MDG Implementation Unit
Mr James M. Mwanzia	Deputy Chief Economist
Ministry of Finance, External Resources Department, GoK	
Mr. Bernard Masiga	Deputy Chief Economist, Head UN Division
Kenya Private Sector Alliance	
Mr. Samuel Mwaura	Chief Executive Officer
Ms. Christine Mwaka	Programme manager
Public Service Reform Programme (PSRP)	
Ms. Mary Njoroge	Head, Programme Management Unit, PSRP
Mr. E. A Lubembe	Programme director PSRP
National Evaluation Consultants	
Ms. Joyce Deloge	Evaluation Consultant
Mr. John Murimi Njoka	Evaluation Consultant
National Civic Education Programme (NCEP)	
Mr. Zein Abubaker	Coordinator
Sida staff	
Ms. Camilla Redner	Senior Programme Officer
Ms. Ulrika Åkesson	Deputy Head
Action Aid	
Mr. David Mwangangi	Policy Manager
Legal Resources Organisation	
Ms. Jedidah Wakonyo Waruhiu	Executive Director
Ms. Alice Njau	Programme Officer
ICJ	
Mr. George Kegoro	Executive Director
United Disabled Empowerment Kenya	
Ms. Salome Kimata	Chief Executive Officer
National Council for Persons with Disabilities	
Representatives	

Uganda Mission 14 - 17 Jan 2008

UNDP Country Office	
Mr. Theophane Nikyema	Resident Representative and UN Resident Coordinator for Uganda
Mr. Sam Jamie Ibanda Igaga	Assistant Resident Representative
Ms. Rose K Ssebatindira	Assistant Resident Representative, former M&E focal point
Ms. Enid Nambuya	Programme Analyst, Governance
Mr. Bharam Namany	HIV/AIDS Program Analyst
Ms. Harriet Karusigarira	Programme officer, Governance
Mr. Justin Ecaat,	Environmental specialist
Mr. Alexander Aboagye,	Economic Adviser
Public Procurement and disposal of Public Assets Authority, GoU	
Mr. Edgar Agaba	Executive Director
Mr. Tumutegyereize G. Milton	Director, Training & Capacity Building
Mr. Kakira N. Geoffrey	Manager, Training & Capacity Building
Office of the Prime Minister (OPM), NEMIS, GoU	
Mr. Sentongo M. Peter	Director NEMIS
Mr. David Rider Smith	Advisor NEMIS, OPM
Ministry of Local Government, Participatory Development Management (PDM) Secretariat, GoU	
Mr. Joseph Kayuza	National Project Director
The Inspectorate General of Government (IGG), GoU	
Ms Jane Kaberuka	Project Coordinator
Planning and development coordination office, Parliament of Uganda, GoU	
Mr. Dison Okumu	Director
Uganda Aids Commission, GoU	
Ms. Rose Mary Nalwadda	Director Planning and Monitoring
The Alliance of Mayors and Municipal Leaders on HIV/AIDS in Africa (AMICAALL)	
Dr. John Mugisa	Director
National Evaluation Consultants	
Mr. James Katorobbo	Evaluation Consultant
Mr. Frank Emmanuel Muhereza	Evaluation Consultant
Ms. L. Keene Mugerwa	Evaluation Consultant

Nepal Mission 21-30 January 2008

UNDP Country Office	
Ms. Anne-Isabelle Degryse-Blateau	Country Director
Mr. Ghulam Isaczai	Deputy Resident Representative
Mr. Sharad Neupane	Assistant Resident Representative, Governance
Ms. Heather Bryant	Monitoring and Evaluation and Knowledge Management Officer
Mr. Dharma Swarnakar	M&E analyst

National Planning Commission, GoN	
Dr. Jagadish Chandra Pokharel	Vice-Chairman
Mr. Subarna Lal Shrestha	Joint Secretary
Ms. Mandira Poudyal	Programme Director
Ministry of Local Development, GoN	
Mr. Ganga Datta Awasthi	Acting Secretary
Mr. Bishnu Nath Sharma	Joint Secretary
Mr. Som Lal Subedi	Joint Secretary
Mr. Mukunda Raj Prakash,	Under Secretary, M&E Section
Supreme Court, Nepal	
Mr. Ram Krishna Timalsena	Registrar
National Human Rights Commission (NHRC)	
Ms. Padma Mathema	Secretary, National Project Director for Capacity Development for NHRC
Mr. Keith D. Leslie	Senior Human Rights Adviser, Capacity Development
Mr. Tek Tamrakar	National Project Manager, Capacity Development
Mr. Kedar Nath Upadhyay	Chairperson of the Commission
Dr. K.B. Rokaya	Commissioner
Dr. Gauri Pradhan	Commissioner
National Evaluation Consultants	
Ms. Kanta Singh	National Evaluation Consultant
Mr. Hemang Sharma	National Evaluation Consultant
Ms. Bikash Sharma	National Evaluation Consultant
Association of District Development Committees of Nepal (ADDCN)	
Mr. Krishna Prasad Sapkota	Chairman ADDCN
Mr. Hem Raj Lamichane,	Executive General Secretary (ADDCN)

Informants

UNDP Evaluation Office	
Ms. Saraswathi Menon	Director
Mr. Nurul Alam	Deputy Director
Mr. Suppiramaniam Nanthikesan	Senior Evaluation Advisor
Mr. Juha Uitto	Senior Evaluation Advisor
Mr. Sergio Lenci	Evaluation Specialist
Mr. Oscar Garcia	Evaluation Advisor
Ms. Azusa Kubota	Evaluation Analyst
Mr. Anish Pradhan	IT Specialist
UNDP Nordic Office	
Mr. Claes Waldenström	Liaison Officer