

**PEER REVIEW OF EVALUATION IN MULTILATERAL OR-
GANISATIONS**

PEER REVIEW OF UNDP EVALUATION OFFICE

KEY ISSUES AND LESSONS IDENTIFIED

17 March 2006

CONTENTS

1. INTRODUCTION.....	2
2. RATIONALE FOR THE ‘NEW APPROACH’	3
3. DESIGN	5
4. STEPS, WORK PROGRAMME AND SCHEDULE.....	5
5. RESOURCES REQUIRED	7
6. KEY ISSUES AND LESSONS.....	8
Demand and Use	8
Panel Composition and Roles	9
Working Relationships with UNDP	10
The Normative Framework and Methodology for Assessment	11
Key Boundary Issues	12
6. OVERALL CONCLUSIONS	14

1. INTRODUCTION

1. At the Development Assistance Committee (DAC) Evaluation Network meeting in January 2004, Denmark tabled a proposal for a 'New Approach' to assessment of multilateral organisations' performance, based on systematic use of the organisations' own evaluations. The proposal was a response to the perceived increase in the demand for multi-donor evaluations of multilateral agencies performance, reflecting increased accountability demands from donors' parliaments and auditor-generals. Responding to this rising demand through more multi-donor evaluations would entail high transaction costs on all the participating organisations and involved partners. It would also fail to recognise the evaluation capacity and autonomy of the multilateral agencies themselves.

2. In response to a positive reception to the proposal at the January 2004 Evaluation Network meeting, the UNDP Evaluation Office (EO) volunteered to participate in the exercise. Denmark then engaged two consultants to: i) examine donor needs for information; and ii) further develop the proposal and a methodology for the 'New Approach'. The report on donor needs was prepared between May and September 2004. The 'New Approach' and methodology were developed, and partially tested with UNDP between June and October 2004. The report on donor needs and a preliminary version of the 'New Approach' proposal were then discussed at a workshop of the DAC Evaluation Network, organised by Denmark, on 23rd November 2004. This workshop included representation from both interested bilateral and multilateral organisations. It reviewed the findings of the donor report and discussed the proposal for the 'New Approach', clarifying a number of issues related to the proposal.

3. The conclusions of the November 2004 workshop and a follow up meeting early April 2005 were reflected in the paper "A New Approach to Assessing Multilateral Organisations' Evaluation Performance, Approach and Methodology, Final Draft, June 2005". This paper outlines the scope of the task, which is to assess the evaluation performance of multilateral organisations and it describes both a proposed approach and methodology for assessing the performance of a multilateral agency's evaluation system. It was also agreed that it should be considered a living document refining it as experience is gained.

4. The New Approach is based on assessment by a panel of professional evaluators. It stipulates that this "peer panel" should comprise representatives from both bilateral and multilateral agencies, and beneficiary countries as well as independent evaluation expertise. This is thus a peer review by professionals, and not a peer review between organisations. The initial focus of the 'New Approach' is akin to a certification of the quality of a particular function within the concerned multilateral.

5. UNDP agreed to undergo the first peer review and the peer panel comprised representatives from Danida, DFID, IFAD, The Netherlands Ministry of Foreign Affairs and IDEAS. The review was conducted during the period September – December 2005.

6. Because it was a first test, the team involved agreed, together with UNDP, to pay special attention to documenting and reporting on the experience and lessons learned for similar exercises in future. This brief report sums up the reflections of the team and the UNDP EO on these points. It is primarily designed for use by Members of the DAC Network and the UN Evaluation Group, and particularly for those who may be planning further assessments in future. While the team stops short here of recommending any major changes in the working approach and methodology for these

assessments on the basis of this single experience, it points to a number of lessons that would suggest some significant changes be considered. A reading of the Review Report, or at least its Executive Summary, is strongly recommended to provide a direct sense of the substance and impact of the work.

7. The experience outlined here should have relevance for applying this approach to any multi-lateral organisation, and particularly to U.N. organisations – since it worked with the new 2005 Norms and Standards for Evaluation in the UN System, and the results of an evaluation self-assessment carried out by all UN organisations. The results of the UNDP Review will be followed shortly after by a review of UNICEF – a different type of UN agency – and the combined reading of the two experiences should prove especially instructive.¹

2. RATIONALE FOR THE ‘NEW APPROACH’

8. Bilateral donors face increasing demands for information on the performance of the multi-lateral agencies that they support. Discussions with staff in five bilateral agencies² suggest that the demand for independent and credible evidence of the effectiveness of the multilateral agencies comes from both the agencies’ own operational staff, and increasingly from their governments’ finance ministries, auditor-generals and parliaments.³ This reflects both an increasing results focus within the bilateral governments overall and, in some cases, a desire to find parameters to help decide how much of ever scarcer aid resources should be channelled through the agencies’ bilateral programmes or through the multilateral agencies, respectively.

9. Reflecting the harmonisation agenda, most bilateral agencies state that they want to rely on performance evidence provided by the multilateral agencies to meet their own growing information needs. Present practice suggests that this aspiration has been only partly achieved. For example, the Department for International Development (DFID), the Danish International Development Assistance (Danida) and the United States Agency for International Development (USAID) increasingly rely upon evidence from multilateral organisations’ performance management and reporting systems for tracking multilateral agencies’ performance at an aggregate level. However, present multilateral reporting systems rarely provide independent evidence of a multilateral organisation’s contribution, or how it was achieved, at country level. This is a key, and growing, focus for the bilateral donors, and evidence suggests that this is a significant driving force underpinning the growing demand for multi-donor evaluations of overall multilateral agency performance as well as parallel initiatives, such as the Multilateral Organisations’ Performance Assessment Network (MOPAN).

10. While bilateral evaluation offices often have limited capacity and are burdened by these resource intensive multi-donor evaluations, some of the multilateral agencies have gradually invested in and built up strong RBM systems and evaluation departments. For example, within the United Nations (UN) system, this has been a key component of the UN reform programme that was launched in 1997, and has also been a response to direction from the governing boards of the agen-

¹ The lead member for the UNICEF Peer Review team was able to join the UNDP assessment team for some of its key meetings and all its materials were shared with her and the consultants for that team.

² See ‘An Assessment of Selected Bilateral Donors’ Current and Future Performance Information Needs’ by Dorte Kabell & Paul Balogun. Danida, November 2004.

³ In the cases of the Netherlands and Sweden there are indications that in future the need for evaluations to meet accountability requirements may increase.

cies. Therefore a ‘New Approach’ that would build on this multilateral capacity, and help reduce transaction costs, while recognizing the accountability requirements of bilateral agencies, is needed. The increased focus on demonstrating aid and development effectiveness – also at the level of multilateral organisations’ governing bodies – will help strengthen these efforts.

11. Some 10 joint evaluations of multilateral agencies have been carried out over the last 10 years. For bilateral agencies, one of the most problematic aspects of multi-donor evaluations of multilateral organisations has been the limited coverage due to the high costs. Therefore, an immediate benefit of the ‘New Approach’ to bilateral evaluation offices would be increased evaluation coverage across the multilateral system. Increased coverage would be possible, because of the increased use of the multilateral organisations’ own evaluations, thus freeing time and resources otherwise invested in collecting primary data. At the same time, transaction costs would be reduced through reducing the need to invest significant time in ensuring the quality of multi-donor evaluations. This also would have pay-offs in terms of the cost-effectiveness of evaluation of the multilateral system overall.

12. For bilateral agencies’ operational staff, the benefits of the ‘New Approach’ would lie in the potential to increase the level of independent and credible evidence of performance available; it is unlikely that the present approach – without a significantly increased number of multi-donor evaluations – would be able to meet this demand. The ‘New Approach’ would address this issue by providing evidence of the independence and credibility of multilateral organisations’ evaluations, hence enabling a wider use of the evaluations from the multilateral organizations. The present practice on the use of evaluation evidence suggests, however, that to significantly benefit operational staff, the ‘New Approach’ also needs to address the actual *use* of evaluation evidence by these staff.

13. However, the rationale for the ‘New Approach’ is not only for bilateral evaluation offices to simply judge the performance of multilateral evaluation offices, in order to shift responsibility for meeting a growing demand. Instead it is a tool that is intended to facilitate the wider movement to a scenario in which bilateral agencies make increasing use of evidence from the concerned multilateral agencies. In such a case the incentives for the multilateral agencies’ own boards to use evaluation evidence, and ensure its quality, are enhanced. As such, the ‘New Approach’ aims to align evaluation practice with that being developed under the DAC-OECD Working Party on Aid Effectiveness and Donor Practices’ Joint Venture on Managing for Development Results, which focuses on performance management and reporting, and is aiming to build capacity and knowledge of ‘best practice’ across all participating organisations. The ‘New Approach’ holds potential for supporting a similar approach by allowing both multilateral and bilateral evaluation professionals to learn from each others’ experience, and potentially establish benchmarks for assessing evaluation performance in the longer term.

14. Such joint learning and confidence building would also benefit the multilateral organisations, as it would help increase evaluation capacity. Furthermore, multilateral organisations would benefit because bilateral donors’ operational staff would, with the ‘New Approach’, provide stronger support for the evaluation functions within the multilateral agencies. The approach contains strong incentives for donors’ board members to make more systematic use of the agencies’ own evaluation evidence. It is generally acknowledged that improvement is needed in this regard.

3. DESIGN

15. The 'New Approach' relies on evaluative evidence from centrally managed evaluation systems of multilateral organisations when assessing its development performance. In case of decentralised evaluation systems the control management in place at central level to ensure adequate quality of evaluation products is to be part of the assessment. The central question is *whether a multilateral agency's own central evaluation office produces evaluations, which are credible, valid and useable for learning and accountability purposes as tested by internationally recognised evaluation peers.*

16. Assessing this requires focusing on three aspects: i) the credibility of the evaluation process and of the evaluation reports; ii) the independence of a multilateral organisation's evaluations and evaluation systems; and iii) the use of evaluation evidence by the multilateral organisations, beneficiary countries and bilateral donors. These can be assessed in two broad ways: by assessing the quality of evaluation processes aimed at delivering independent and credible evaluations and ensuring the use of evaluative evidence and by directly assessing the quality of content of actual evaluations produced and if evaluative evidence is actually used.

17. The methodology was based on developing a series of normative assessment frameworks for the internal consistency and effectiveness of evaluation systems and policies and the use of evaluation evidence; and the independence of evaluations and evaluation systems; and the credibility of evaluations. Making these assessments require a set of norms and standards. In the absence of recent, all encompassing and internationally agreed standards, it was originally foreseen to use the DAC standards for these assessments, as well as the norms and standards used by individual agencies. However, as the UN Evaluation Group⁴ (UNEG) completed in April 2005 a set of norms and standards for evaluation, it was agreed to use these as the basis for the assessment.

18. The Panel directly applied the set of thirteen broad Norms for Evaluation in the UN System, amplified by the more detailed UNEG Standards. These had first been cross-checked against the relevant issues in the DAC Principles. Not surprisingly, given the acknowledged influence of the DAC Principles for evaluation in shaping the UNEG Norms and Standards, it was found that they fully reflected almost all the Principles.

4. STEPS, WORK PROGRAMME AND SCHEDULE

19. The broad steps involved in carrying out the assessment followed the sequence set out in the Approach and Methodology as outlined below. One variation was that the work was able to take advantage of a good deal of the preliminary familiarization with UNDP, data-collection, and analysis carried out during the exploratory period in mid-2004. An important practical consideration was that the assessment was designed to be targeted, and to minimize additional or duplicative workload demands on UNDP personnel, either in the Evaluation Office (EO) or elsewhere:

⁴ The group has 34 member evaluation offices and its membership covers both the UN organisations and the multilateral banks.

- Step 1: Development and agreement on the basic normative frameworks for assessing independence, credibility and use of evaluation evidence;
- Step 2: Collection of data, and its analysis, against these normative frameworks;
- Step 3: Agreement reached on the accuracy of the evidence and findings;
- Step 4: Development of conclusions and recommendations.

The more detailed work-programme for this particular Review evolved as follows:

20. Building on the earlier groundwork on UNDP, the Review got fully underway with the first meeting of most members of the panel and their advisors to agree on the work-plan and the normative frameworks to be proposed. (Week 2: End of August 2005)

21. Reviewing the preliminary assessment from November 2004, already fact-checked with UNDP EO, to transpose, crosscheck, up-date, and fill gaps as needed to reflect the proposed assessment framework and changes at the UNDP end. A new draft Evaluation Policy for UNDP, as well as a self-assessment “quality stamp” questionnaire sent to all UN agencies by the UN Evaluation Group earlier in 2005 and the self-assessment responses provided by UNDP in September were taken fully into account as very valuable new inputs to the assessment design, preparations and information base. (Weeks 1-5).

22. The normative frameworks, further data collection, interview strategies, milestones and other issues were discussed and agreed upon with UNDP’s Evaluation Office (EO) (Week 8).

23. Extensive review and updating of the documentary and website information base gathered in 2004, supplemented by additional draft and other materials supplied by UNDP EO in mid-October 2005 meetings. This base proved to be extensive and useable, and UNDP EO continued to offer and provide further materials as required; (Ongoing from week 1 to week 14).

24. To make the informational foundations as concrete, focused and systemic as possible within the scope this review, a small group of recent evaluation processes and products of different types was selected by the Panel, in close collaboration with UNDP EO, as reference cases taking into account a combination of several criteria.⁵ These were not expected to be full process or product assessments. The reference cases provided a basis for more detailed information gathering by the Panel’s Advisors, including structured and semi-structured interviews with internal and external participants in these processes from different perspectives. The evidence was fed into the compilation of findings, with some preliminary triangulation among the various sources and types of information; (Week 8).

25. Advisors provided the Panel with the information base to date, specifying the sources, and organised under Policies to date, Practices, and Future Directions (with the latter reflecting draft policy and innovations now being tested). This information was inserted into the existing Proposed Framework, and it did not go beyond findings at this stage. This compilation was provided to UNDP EO for their information and any rapid fact checking and feedback needed prior to the panel interviews, and for later refinement if necessary in the draft report; (Weeks 8-9).

⁵ These featured the need for the examples to be relatively recent, relevant to the main types of UNDP work and evaluations, at different stages of follow-up, and with reasonable accessibility of knowledgeable informants.

26. Peer interviews: With the benefit of the information assembled, and its individual and collective examination by the panel, including selected primary sources, and any observations to date from UNDP EO, the Panel conducted wider interviews with a number of UNDP Senior Managers and Executive Board members. The reference cases served as one important focus; (Week 10).
27. Integrating the results and insights from these interviews, the Panel completed the triangulation, refinement and confirmation of its base of evidence and findings. On that basis, it proceeded into the “judgement phase” in these steps: panel agrees on its main frameworks for judgements in relation to the Norms and main aspects of the review; panel considers and debates the evidence and findings, and arrives at its draft conclusions and recommendations, agrees on draft report; (Week 13).
28. Draft report serves as the basis for the Peer Review meeting with UNDP EO to consider the results; (Week 14).
29. Panel carries out final review and any necessary revision to the draft report; (Week 14).
30. Panel transmits its final report to UNDP EO, providing final opportunity for review and reflecting dissenting views or other key responses; (Week 15).
31. Panel presents Peer Review Report to UNDP; (Week 16).
32. The Administrator provides UNDP’s response to the Review, to be posted on the website and incorporated with the Peer Panel’s report; (Week 24).
33. Panel Chair to present report to an informal meeting of the UNDP Executive Board, in tandem with the presentation of the Draft Evaluation Policy for the organisation; (Week 28).
34. Panel transmits its report to the DAC Evaluation Network and UNEG, together with its report on lessons learned from the exercise; (Weeks 29-31).

5. RESOURCES REQUIRED

35. The main resources expended by the team during the review involved a total of 250 professional workdays, comprising 130 for five reviewers and 120 for two consultant advisors. It should be noted that the consultant time required was somewhat reduced by the exploratory work carried out prior to the summer of 2005, which is not counted here. Travel and other costs for the full team amounted to a total of some US \$125,000.
36. As the host organisation, and especially as the partner in the first such review, the UNDP’s Evaluation Office contributed sizeable amounts of substantive and support time to the exercise. Beyond the Evaluation Office, some 60 other UNDP officials and 20 Executive Board members were involved in setting up, preparing for and undertaking individual interviews and meetings during the review process, and a number (including senior level officials) were involved in launching, concluding and following up the review.

6. KEY ISSUES AND LESSONS

Demand and Use

37. The 'New Approach' to assessing multilateral organizations' performance was developed in direct response to the evolving practice of labour intensive and expensive multi-donor evaluations of multilateral organizations. The approach, it is believed, is a viable and cost-effective alternative, that should also have the benefit of improving confidence in the use of multilateral agencies' evaluations by a broader community, including bilateral agencies themselves.

38. This hypothesis was tested through an assessment of five bilateral donors current and future performance information needs, which itself was fed into the development of the 'New Approach' itself. This assessment found that bilateral information needs are multifarious, and that evaluations in general (whether produced by bilateral or multilateral agencies) are not always used to their greatest potential.

39. The expressed level of demand and potential use of the findings of the Review from within the multilateral organization itself, the Evaluation Office (EO), the UNDP management, and the UNDP Executive Board varied. The EO, as volunteers, and the UNDP management recognized the value of the Review, supporting as it did the simultaneous development of the UNDP Evaluation Policy. The UNDP Executive Board was not consulted during the design phase of the 'New Approach', and it proved difficult to formally engage it in the decision to perform the review. The Director of the EO informed the EB of the process in June 2004, again its meeting in June 2005, and informal briefings and interviews were arranged with interested Board Members as it progressed, with the cooperation of the Bureau. The culmination of this linkage was the presentation of the panel's report by the panel Chair at the time of the Executive Board's informal discussion in March 2006 on the proposed new evaluation policy for UNDP. The lack of formal engagement with the EB did not allow for the desirable degree of ownership by the board in this case. To underline their direct stake and the key importance of the accountability function, the Panel's Final Report explicitly highlighted a number of suggestions addressed to the Executive Board and Administrator for their consideration and possible action, as well as suggestions to the Evaluation Office itself. Fortunately, the coincidence of the peer review with the drafting and discussion of the new evaluation policy has ensured that the EB did make use of the review, as evidenced in the comments and suggestions made at the presentation.

Lessons:

- **Based on the principal that the use of any assessment is contingent on establishing a clear demand and use up-front, considerable effort is necessary to clarify and engender participation within the donor community, and with the UNDP Executive Board. To help strengthen the engagement and knowledge of members of multilateral governing bodies in strengthening and using evaluation, ways should be sought to formally engage the governing body of the entity being reviewed at an early stage, even if this may take more time. The suggestion in the June 2005 Approach paper, of seeking a Board Member to sit on the Peer Panel was found politically impractical. To build greater governing body engagement in future reviews there may also be a special need to demonstrate that these exercises are undertaken in the shared interest of all Member Countries and their citizens and not just a group of donors who took the first initiative.**

- **Beyond the Review’s value to the UNDP EO and management, its use will only be evident in the counterfactual, namely by preventing the conduct of a multi-donor evaluation of UNDP, and through evidence of the use of specific evaluations by bilateral departments. To determine this, it may be helpful to conduct a study of actual use at some later stage.**
- **In view of the ‘New Approach’s’ primary objective of replacement for an institutional evaluation, future Reviews should also assess the coverage of evaluations conducted by the central evaluation unit. While UNDP’s EO has evaluated a number of key organizational dimensions over the past three years, and thus has sufficient coverage, this was not an explicit criterion in the Review.**

Panel Composition and Roles

40. The prime initiative – and convening and coordinating responsibility – for this exercise rested primarily with the “lead donor,” the Head of Evaluation Department of the Danish Foreign Ministry, who served as the Panel Chair and spokesman, in a collegial mode of operation. In consultation with the other designated agencies from the DAC group (the Netherlands and UK evaluation offices) he was responsible for recruiting the panel members. In this case, two of the bilateral evaluation offices designated senior staff members as their representatives, while the third designated a retired senior manager from the programme side (bringing the perspective of an experienced evaluation “user”). The head of evaluation for another UN organisation, IFAD, also agreed to serve, as did an independent evaluation expert from an international development evaluation association (also bringing the perspective from a “non-donor” country). While these Panel members were appointed by their respective organisations, they all served in their capacity as professional peers, bringing their individual judgement to the process and the report and their conclusions were not subject to “clearing” by their organisations.

41. Two senior consultant advisors were selected to assist the Panel with a good deal of the preliminary work. Together, they brought substantial comparative experience in development cooperation and evaluation, together with knowledge of UNDP (from the preceding formative work) and special background in peer review work.

Lessons:

- **These reviews should expect to encounter challenges in identifying appropriate panel members from among other multilateral donor agencies, beneficiary countries, and independent evaluation experts. They may have to handle potential real or perceived conflicts of interest between certain agencies, or between consultants and the host agency as a client, and reconcile the experts’ availability for this work with their normal workloads. One question that the lead agency should consider in advance is what constitution of peers those within their agencies that are commissioning multi-donor evaluations of multilateral agencies will deem credible.**
- **There are important arguments for seeking early and structured participation in future reviews from a wider group of countries, and the interest of UNEG in related work could be an asset in this. The more active role and engagement of UNEG in this work has been positive development. Perhaps there is an analogy to the broader dynamics around the international development strategy, with much greater joint activity in all aspects of development work, from needs identification all the way to evaluation. There could be consid-**

erable benefits to all concerned if ways could be found for future reviews to include more peer panellists and advisors from developing countries, with the active support of developing country members of the governing body of the organisation concerned.

- **The Panel found that the diverse backgrounds and experiences of its members and advisers was a strong combination that brought out varied perspectives and insights. The innovations of incorporating an experienced “user” of evaluation and an evaluator from beyond the development cooperation field probably had the effect of broadening the analysis from what otherwise might have been a narrower, “in-group” exercise and attuned the treatment better to the non-specialised audiences for whom it was mainly intended. All agreed that the gender balance should have been better. UNDP EO, moreover, believes that there was uneven knowledge of UNDP among the members of the panel and that the Panel should have had stronger representation from a multilateral agency with country presence. These different views suggest that all concerned be clear on their primary audiences.**
- **The kind of streamlining of roles done in this review made sense, while still respecting the basic requirements of the approach. Energetic, experienced and diplomatic consultant support is likely to be essential. The Panel should be substantially engaged from an early stage, to provide overall directions, key issues, questions and possible hypotheses. Consultants can then pursue the issues and gather evidence in greater depth, then make a clear “handover” of evidence and findings to the Panel, allowing the time and access to sources to test their accuracy and proceed to forming judgments, conclusions and possible suggestions.**

Working Relationships with UNDP

42. Linkages with UNDP’s Evaluation Office were designed to be close and collaborative throughout the process, as befits both a peer review and the collegial development of a new approach. Both the Evaluation Office and the Panel and its advisors were sensitive to the Panel’s need for independence in the review work as well as to all reasonable requirements for access to information and well-organised consultation with others in UNDP. The Panel and EO worked together to ensure that the review would be seen as relevant, important, and constructive for the Organisation’s work; not duplicating or demanding too much time and effort of EO and/or others in UNDP for unclear benefits; and not being seen to give undue attention or access to any special concerns of a group of donor countries.

43. Once the ground-rules had been clearly worked out together – which needed to be done in some detail – the working relationships with the UNDP Evaluation Office proved to be productive and mutually informative. It was noteworthy that at the draft report stage the EO requested the Panel to provide more specific examples and suggestions of good practices elsewhere that might be suitable models for UNDP to adopt or adapt. The Panel stopped short of attempting to prescribe at that level, making generic suggestions drawn from its collective knowledge and experience, while leaving it to UNDP to come up with the approach best suited to its particular circumstances.

44. A briefing letter was circulated early on to introduce the project and the “peer review” approach to UNDP board, staff, and prospective interviewees. From the outset and at key milestones, the panel worked with the Director of the EO to ensure that, together with other senior managers,

the Administrator of UNDP was personally well-informed of this work, both in his capacity as the head of management and in his reporting relationship to the Executive Board as governing body. He and many other senior managers also gave direct interviews on their own views, needs and expectations of the evaluation function, and the Administrator provided the Organisation's response to the report. Especially in view of the transition in Administrators at UNDP in the midst of this review, senior management provided its full cooperation, encouraged by the openness of both the outgoing and incoming chief executives. This openness seemed to have been supported by the work done to explain and situate the character, style and benefits of peer review, bearing in mind that some aspects that are now taken for granted among DAC donor countries. The fact that a number of senior UNDP managers had also been consulted during the exploratory phase may have contributed to some sense of overload on their part, which it should be possible to avoid in future with the more developed methodology now in place.

Lessons:

- **It was suggested by the Evaluation Office that future panels should invest even more heavily in preparatory work in order to best clarify their information needs and especially interview requests.**
- **The Panel concluded that at least one earlier and more intensive peer panel meeting would have been helpful, although it would be very difficult to replicate more rapidly the strong base of information and understanding of the Organisation accumulated during the exploratory work a year earlier when the peer approach had not yet been selected.**

The Normative Framework and Methodology for Assessment

45. The three main areas for these assessments are independence, credibility and utility. A good number of the UN Norms and Standards cross-cut more than one of the three (not surprisingly since these areas themselves overlap) but it was broadly possible to group most of them under the three headings, applying judgments and interpretations that did not have major substantive implications. In spite of the occasional overlaps and repetitions in using these three main headings, they stood up reasonably well as the basis for a framework of assessment, and sometimes served as a useful reminder of the qualities required of evaluation, especially to the non-evaluators consulted. The Standards were expected to help suggest the questions to use in checking the implementation of the Norms, but their detail was found at this stage to be somewhat lopsided,⁶ so that the Panel questions in some areas had to be derived directly from the Norms. UNDP's existing guidelines for evaluation provided a further important reference for information and assessment. The Framework was used directly to assemble a very large compilation of information and findings as they were accumulated. Maintaining this single working document helped ensure thoroughness and transparency in the base of evidence⁷.

⁶ The huge preponderance of coverage in the Standards is on "Credibility" and management of evaluation systems, and 18 of the 49 Standards are directed specifically, and in considerable detail, to Evaluation Reports as such. On "Independence" issues, there is reasonable formal coverage, although since they are not rules-based, their application in practice seems to be heavily dependent on custom, tolerance or support by Heads of Organisations and/or interest and support by their Governing Bodies.

⁷ The document summarizing this compilation (90 pages) is available on request from UNDP Evaluation Office: contact.eo@undp.org, and from the Danida Evaluation Department: eval@um.dk.

46. As the work evolved it was necessary at several points to elaborate explicitly and clarify the methodology being applied. While the June 2005 approach paper outlines the scope of the task and describes both a proposed approach and methodology, the Panel and EO still needed to agree on exactly how the framework would be applied. The clarifications centred on an operational definition linked to the UNEG Norms and Standards and on how the fact that the review itself would necessarily be less comprehensive and in-depth than a formal evaluation, but as outlined in the final report, adhered to a rigorous methodology, applying the key principles of evaluation while taking full advantage of the particular benefits of a peer mechanism. In introducing the exercise to interlocutors in UNDP and eliciting their cooperation, the peer review approach was specifically described and illustrated.⁸ It was made explicit that “the final conclusions will clearly be a judgment” by the Panel.

Lessons:

- **The justification for establishing a framework beyond taking the UNEG Norms and Standards as drafted was based on the need to establish a common basis which both the Peer Panel and the multilateral agency agree. This principal of a ‘negotiated agreement’ will be important for future Peer Reviews where the UNEG Norms and Standards may not be the foundation (such as the MDBs, where good practice standards are used).**
- **The normative framework adopted covered the ground well. The compilation of evidence and findings in a single working document (structured around that framework) permitted the systematic and transparent organisation of findings, and provided a basis for effective consideration of conclusions, suggestions, and a clearly drafted report.**
- **Drawing on the clarification of the “New Approach” methodology from the pilot experiences, the process of future reviews should begin with a discussion and clear agreement by all parties on a brief outline of the framework and methodology. This should serve to clarify expectations and provide a sound basis for agreement on a more detailed review plan and work-programme.**

Key Boundary Issues

47. Three other important issues relating to the scope and limits of this exercise had been anticipated from the outset in the new approach and needed to be resolved, or at least managed, for this review to proceed effectively.

Direct and Indirect Assessment

48. The first of these questions was the extent to which the exercise would attempt to assess directly the quality of content of actual evaluations produced and how much evaluative evidence is actually used. For a number of practical and methodological reasons,⁹ the suggested methodology in June 2005 instead favoured a more indirect approach, focusing on the quality of evaluation processes aimed at delivering independent and credible evaluations and ensuring the use of evaluative

⁸ For example, the Panel arranged to circulate relevant excerpts from Fabiano Pagani, “Peer Review: A Tool for Cooperation and Change: An Analysis of an OECD Working Method.” OECD, September 2002.

⁹ Especially the absence (to date) of firmly agreed standards for assessing quality of content of actual evaluations produced or the use of evaluative evidence, and the time, resources and access that would be required.

evidence. This assumes that there is a strong relationship between the quality of process and the quality of the product or outcome.¹⁰

49. As this review unfolded in practice, the Peer Panel developed something of a mixed approach on this point. The first concern was to assess the capacity of UNDP's evaluation systems to produce at the expected standards of independence, credibility and utility. But in order to actually arrive at this assessment, it proved both necessary and possible to identify and analyse a set of broadly representative "reference cases" of different types of recent evaluation in order to ground and/or validate the panel's findings on the systems. Fortunately, the UNDP EO's evaluative base allowed for a relatively easy and uncontroversial selection and the panel found useable working criteria for making some assessments of both quality of content and the use of evaluation.¹¹ The results of this exercise may provide a useful source for further work underway in these areas.

Evaluation and Other Systems

50. While planning this process to focus on the evaluation function, and not attempting any comprehensive institutional assessment, all concerned were aware that evaluation does not operate – in any organization – in isolation from other systems for management and accountability, such as the system for results-based management, internal audit, etc. In practice, it quickly became apparent that managing the boundaries of the exercise was going to be a difficult challenge, particularly in relation to UNDP's RBM systems.

51. In relation to evaluation in particular, resources available and time schedules dictate that EO managed evaluations must rely and build to a large extent on information that is already available. In terms of identifying UNDP's results and contribution, this means relying upon information available through the RBM systems. Meanwhile, the evaluation function, unlike in some organisations, is not mandated to validate directly the self-assessments of performance against the strategic results framework (the Multi-Year Funding Framework) used in UNDP's RBM system.

52. The Peer Panel had to come to grips with how to deal with the fact that while UNDP's RBM systems are not within the control of the EO, they have major implications for its work. The issue was managed, if not resolved, by taking into account the substantial evidence and findings that emerged on RBM in relation to evaluation and overall accountability in both the Panel's deliberations, report and suggestions to Senior Management.

Centralised and De-Centralised Evaluations

53. Another key boundary concern raised similar difficult issues. As in many other development cooperation organisations – multilateral and bilateral – the central evaluation office in UNDP accounts for only a small share of the total evaluation output, albeit the majority of the work that is

¹⁰ As a related example, in spite of its title, the UNEG Task Force's self-assessment "checklist on evaluation quality" does not clearly lay the ground for any assessment of "quality" but more of the levels of formal conformity with the UNEG Norms & Standards as a presumed basis for quality in evaluation. Even then, it is silent on quality control systems themselves.

¹¹ Among the sources, the very detailed treatment in the 18 UNEG Standards on "Reports" does provide an extensive set of objective and other indicators for assessing the "quality" of actual evaluations produced, and these are very similar to working documents being tested in the DAC Evaluation Network. UNDP EO's own thinking and draft materials on ways of strengthening quality standards was also helpful. Although there is a similar lack of hard, agreed indicators of the "use" of evaluative evidence, the door is opened wide by many of the UNEG Norms & Standards, and the 1998 DAC Review questions for evaluation users suggested some robust and reasonably objective criteria for eliciting informed judgments from intended evaluation users themselves.

cast at a strategic level. Meanwhile, the lion's share of individual evaluations is produced or sponsored by operational departments and country offices, with the central office providing some level of guidance, support, quality assurance and quality control. This de-centralization is usually considered a normal and healthy arrangement for this type of organisation, but it always presents challenges in tracking and assuring the quality of the whole range of evaluation activities. For this reason, the June 2005 methodology specified that the multilateral assessments would focus on the production of the "multilateral agency's own central evaluation office."

54. However, it is also true that, in addition to their intrinsic importance for operational managers, these "decentralised" evaluations (managed by country offices and bureaux) on the actual outcomes of UNDP's projects and other activities are among the indispensable "building blocks" for a strategic evaluation function to serve properly the needs of the Organisation as a whole, as well as a key foundation of its evaluation practice and culture. It should be noted that the UNEG norms and standards for evaluation are clearly concerned with the whole range of evaluation activity in the Organisation.

55. Thus, as reflected in the title of the final report, this review focused on the work of the central evaluation office – but taking into account its relationships with the many evaluations carried out elsewhere in the organisation. Despite this, through interviews, inference and corroboration by direct observation of some of the reference cases; the Review did recommend some measures for improvement in quality assurance and quality control. It is also noteworthy that a positive response to these suggestions was also a key element in the Administrator's response.

Lesson:

- **Reviews of the evaluation function will need to anticipate and manage important "boundary issues" relating to the lines between assessing evaluation processes and products; centralized and de-centralized evaluation systems; and between evaluation and other systems for performance information and management, such as RBM. The approach of maintaining a central focus, but explicitly taking account of the essential linkages proved reasonably successful in this case, making quite clear the extent and limits of the validity of its findings beyond the main boundaries. In the case of the UNDP review, the agreed decision not to cover evaluation in the Funds and Programmes was in retrospect a shortcoming, although to do so would have considerably increased the work and complexity involved.**

6. OVERALL CONCLUSIONS

56. Overall, the 'New Approach' and its first application may be considered successful. Through the Review, the evaluation peers of UNDP's EO were able to come to reasonably confident conclusions regarding the core question of the quality of evaluations produced. These conclusions included a positive assurance on their value for learning and strategy formulation in UNDP, targeted suggestions in each of the areas of independence, credibility and utility of evaluation and a number of suggestions to strengthen their value for accountability, of which some of the key ones appear likely and to provide.

57. Even while the Peer Review was still underway, UNDP's EO indicated that it was already having a beneficial impact in supporting and making inputs to the work that they and others in the UN system were doing to strengthen evaluation. For UNDP in particular, the recognition of the need to set out a state-of-the-art evaluation policy, and many ideas for its content were acknowledged to have been reinforced or suggested by this peer exchange process. As is stated in the Administrator's response (now available with the report) the processes leading to the final peer review meetings and the written report helped encourage several further important management decisions for policies and practices in this area, although the Panel still sees serious concern with the dearth of formal provisions for independence in staffing the post of Director of Evaluation and other posts in the Evaluation Office¹². From this point forward, the key tests will be how effectively the directions set out for policy and practice will be implemented in practice, and how much further engagement is achieved by the Executive Board.

58. On the basis of this evidence, it seems fair to conclude that the review process with UNDP has had useful outcomes that should lead on to further beneficial impacts in helping enhance the Organisation's evaluation capacity and performance. Moreover, in the full spirit of such a peer exchange, the Members of the Review Team and their home institutions have also derived substantial benefits from this in-depth exposure to different ways of handling some important shared challenges. The approach and methodology as applied and adapted in this first case have proved viable, and a number of points for possible modification or flexibility in the June 2005 Approach and Methodology paper have been identified in the preceding sections.

59. As suggested by the description of the process, the expenditure of time, effort and financial resources by all concerned to achieve these results has been considerable.¹³ While this cost is far less, for example, than for the types of independent multi-donor evaluations of multilateral programmes commissioned over recent years, the latter are obviously more far-reaching. The results should be considered in this light.

60. Finally, it remains for each reader of the full Review report and this report on issues and lessons to ask to what extent they might help respond to a major underlying question as to whether Members of the Organisation can have confidence that the central evaluation office itself produces broad evaluations of the Organisation's work that are of comparable quality to those that would be produced if commissioned through independent multi-donor evaluations.

Lessons:

- **While organisations can be expected to vary significantly, the initial investment during 2004 was specific to the piloting nature of this first Review, and thus should not be necessary in future review exercises. Beyond the pilot phase, the effort expended in the review proper was reasonably efficient for this type of task. Comparison with the experience of the UNICEF review should shed further light on this point.**
- **With the type of approach and methodology used here, it is possible for such a peer review to provide a reasonably confident and documented response to the central assessment**

¹² The comments by members of the Board at the informal meeting on 9 March 2006 indicate that the Board will consider this issue.

¹³ In this first case, in particular, considerable additional effort was invested by all in the exploratory phase, while the approach was being developed.

question. The results need to be shared widely with all decision-makers concerned to see to what extent they will share this conclusion, and work more closely with multilateral organizations' own evaluation capacities. This will determine the extent to which such efforts may complement or substitute for wider institutional reviews in future.

- **Each multilateral organization considering participation in such an exercise should make its own assessment as to the likely utility of the process and findings to the agency, in the light of this and other experiences.**
- **The value of this type of exercise for all concerned will be greatly enhanced if it can be planned and scheduled to mesh with important decisions, reviews and/or relevant events within and beyond the agency being assessed.**

ooo0ooo