

Peer Review EVALUATION FUNCTION_{AT} THE WORLD FOOD PROGRAMME_(WFP)

PEER PANEL MEMBERS: Jock Baker, Stefan Dahlgren, Susanne Frueh, Ted Kliest, Zenda Ofir
ADVISORS TO THE PANEL: Ian Christoplos, Peta Sandison



Peer Review
EVALUATION
FUNCTION_{AT}
THE **WORLD FOOD**
PROGRAMME (WFP)

PEER PANEL MEMBERS: Jock Baker, Stefan Dahlgren, Susanne Frueh, Ted Kliest, Zenda Ofir
ADVISORS TO THE PANEL: Ian Christoplos, Peta Sandison

Published by Sida 2007

Authors: Jock Baker, Ian Christoplos, Stefan Dahlgren, Susanne Frueh, Ted Kliest, Zenda Ofir and Peta Sandison.

The views and interpretations expressed in this report are the authors' and do not necessarily reflect those of the organisations commissioning this study.

Graphic Design: Edita Communication AB

Cover: Gaute Hanssen

Article number: SIDA39657en

Printed by Edita, 2007

ISBN 978-91-586-8241-2

This publication can be downloaded/ordered from www.sida.se/publications

Foreword

The present report is the result of the third professional peer review of the evaluation function at a multilateral development or humanitarian organisation. It was carried out at the request by WFP to the DAC Network on Evaluation and the UN Evaluation Group (UNEG). The review is the joint effort by a group of international evaluation experts who, in close collaboration with the evaluation office at WFP, have assessed the strengths and weaknesses of the evaluation function at WFP. A number of measures for improvement are proposed in the report.

The reason for these peer reviews was originally a wish by bilateral donors to rationalise the external evaluations of multilateral organisations by relying more on the multilateral organisations' own evaluations and, by extension, to be assured about the integrity of their evaluations. The larger aim of the review is to stimulate constructive thinking about how to further improve the two main evaluation purposes: *learning* and *accountability*. In the words of our guiding document *Framework for Professional Peer Reviews* established by the joint UNEG-DAC peer review task force:

“This approach has several purposes: building greater knowledge, confidence and use of evaluation systems by management, governing bodies and others; providing a suitable way of “evaluating the evaluators”; sharing good practice, experience and mutual learning. The primary intended audience for the results of these professional peer reviews is one of decision-makers and other users of evaluation – including where appropriate the intended beneficiaries in member countries.”

Following the *Framework* we have assessed the evaluation function along three criteria or dimensions: *independence*, *credibility* and *utility*. The norms and standards for evaluation developed by the UNEG have provided more precise definitions and ways to operationalise these three

criteria. Our conclusions are summarised at the end of each of the three main chapters of the report, where we indicate to what extent the UNEG norms are met.

The review was truly a learning experience for us and we enjoyed a very open and cooperative response from all parts of WFP. We would like to thank everyone who has set aside time for interviews and provided information in other ways, particularly the staff at OEDE who made the review possible by generously sharing facts and comments along the way and by smoothing our path to gain better knowledge about WFP.

Stockholm, November 5th, 2007

For the Peer Panel

Stefan Dahlgren

Chair

Table of contents

Abbreviations and acronyms	7
Executive summary	9
1. Introduction	19
1.1 Background and purpose of the Peer Review	19
1.2 The Review process.....	20
1.3 The Peer Panel members and advisors.....	20
1.4 The focus of the Review	21
1.5 Methodology	23
1.6 Limitations and challenges of the Review.....	25
1.7 Organisation of the report	27
2. Evaluation at WFP	29
2.1 The current evaluation system in WFP	29
2.2 Approach to evaluation.....	33
3. Independence	35
3.1 Role of the Executive Board and Executive Director.....	35
3.2 Independence in programming, carrying out and reporting of evaluations ...	36
3.3 Independence and impartiality of evaluation managers and evaluators	38
3.4 Links between evaluation planning and budget	40
3.5 Relationship between evaluation and audit	41
3.6 Ensuring access to needed information	42
3.7 Conclusions on independence.....	42
4. Credibility.....	45
4.1 Evaluation policy	45
4.2 Monitoring and Evaluation Guidelines	47
4.3 Credibility of data used in evaluations.....	48
4.4 Competency and capacity of leadership, staff and consultants.....	49
4.5 Quality assurance	50
4.6 Stakeholder consultation and engagement.....	51
4.7 Support to decentralised evaluations	51
4.8 Role of self-evaluation and alternatives such as After Action Review	54
4.9 Basic criteria for quality of evaluation.....	55
4.10 Credibility of recommendations.....	57
4.11 Conclusions on credibility	57

5. Utility.....	59
5.1 Purpose of evaluation	60
5.2 Uses of evaluation for programme decision-making	60
5.3 Meeting the needs of different users of evaluations	62
5.4 The management and follow-up of evaluations	64
5.5 Joint and inter-agency real-time evaluation	65
5.6 Contribution of evaluation to managing for results	66
5.7 Contribution of evaluation to policy making and organisational learning	68
5.8 Contribution of evaluation to knowledge management	70
5.9 Management response to evaluations.....	71
5.10 Contribution to building local M&E capacity.....	72
5.11 Conclusions on utility	73
6. Principal recommendations	75
6.1 Relationship between OEDE and the Executive Board.....	75
6.2 Evaluation policy and strategy	76
6.3 OEDE's mandate.....	77
6.4 External accountability; relations with partners and stakeholders	78
6.5 Management response.....	80
6.6 Evaluation quality	81
6.7 Organisational learning.....	82
6.8 Monitoring and RBM.....	83
6.9 Team selection and procurement of external evaluation expertise	83
6.10 Staffing of OEDE	83
6.11 Budget for evaluations	84
Appendix 1 Approach and Work Plan	85
Introduction.....	85
Background	85
Purpose of the Professional Peer Review.....	86
Scope of and Limitations to the Professional Peer Review.....	87
Core assessment questions	88
Normative framework	89
Panel composition	89
Reporting.....	90
Responsibility of OEDE.....	90
Review process.....	90
Schedule	91
Resources	92
Appendix 2 Framework for Professional Peer Reviews	93
Summary	93
Background	95
Rationale	96
Approach and methodology: Recommendations and options	97
Roles and responsibilities in the reviews	104
Phase-by-phase organization of a professional peer review process	107
Key references.....	110
Appendix 3 Persons interviewed	111
Appendix 4 Background data for the Peer Panel members and advisors.....	117

Abbreviations and acronyms

AAR	After Action Review
ALNAP	Active Learning Network for Accountability and Performance in Humanitarian Action
COMPAS	Commodity Movement Processing and Analysis System
DAC	OECD's Development Assistant Committee
DFID	Department for International Development (UK)
DSC	Direct Support Costs
EB	WFP's Executive Board
ECOSOC	UN Economic and Security Council
ED	Executive Director
EMOP	Emergency Operation
EMSYST	Evaluation Memory System
FAO	Food and Agricultural Organisation
IA-RTE	Inter-Agency Real Time Evaluation
IASC	Inter-Agency Standing Committee
IEG	Internal Evaluation Committee
IFAD	International Fund for Agricultural Development
IPDET	International Program for Development Evaluation Training
M&E	Monitoring and Evaluation
NGO	Non-Governmental Organisation
OCHA	UN Office for the Coordination of Humanitarian Affairs
ODA	Analysis, Assessment and Preparedness Division (WFP)
ODS	Office of Oversight, Operational Audit and Inspector General (WFP)
OEDE	WFP's Evaluation Office

OEDR	Results-based Management Division (WFP; now dissolved)
PDP	Policy Strategy and Programme Support Division (WFP)
PQA	Program Quality Assurance Group (WFP)
PRC	Programme Review Committee (WFP)
PRRO	Protracted Relief and Recovery Operation
PRSP	Poverty Reduction Strategy Paper
PSA	Project Support and Administration
RBM	Results-Based Management
RTE	Real Time Evaluation
SENAIP	Strengthening Emergency Needs Assessment Implementation Plan
Sida	Swedish International Development Cooperation Agency
TEC	Tsunami Evaluation Coalition
ToR	Terms of Reference
UNDAF	United Nations Development Assistance Framework
UNDP	United Nations Development Program
UNEG	United Nations Evaluation Group
UNICEF	United Nations Children's Fund
USD	United States Dollar
VAM	Vulnerability Analysis Mapping
WFP	World Food Programme
WINGS II	WFP Information Network and Global System Phase Two

Executive summary

Overview

The purpose of this Professional Peer Review is to provide an independent assessment of the functioning and quality of the WFP evaluation function. It is intended for use by WFP's leadership, WFP's Executive Board and the OEDE, as well as the DAC and United Nations Evaluation Group (UNEG). The findings of the Professional Peer Review are to be presented to WFP's Executive Board meeting in February 2008 and will inform further discussions and decisions about the function and administration of OEDE. The assessment report is also to be provided for information to the DAC/UNEG Task Force and will be made publicly available.

The Professional Peer Review takes the central evaluation function, i.e. OEDE, as its starting point but also includes analysis of decentralised evaluation in WFP and the ways in which the organisation as a whole engages in evaluation. Significant attention is paid to the attitudes and perceptions that frame how WFP uses and learns from evaluation. In line with the *Framework for Professional Peer Reviews of Evaluation Functions in Multilateral Organisations*, established by the DAC/UNEG Task Force, the Peer Review of the Evaluation Function of WFP has applied three core criteria that need to be satisfied for evaluation functions and products to be considered of high quality:

- A. *Independence* of evaluations and the evaluation system(s). The evaluation process should be impartial and independent in its function from the process of policy making, and the delivery and the management of assistance. Independence of the evaluation function is a pre-condition for credibility, validity and usefulness.
- B. *Credibility* of evaluations. The credibility of evaluation depends on the expertise and independence of the evaluation managers and the evaluators and the degree of transparency of the evaluation process.

Credibility requires that evaluations should report successes as well as failures. Partner countries and representatives of other partners should, as a rule, fully participate in evaluation in order to promote credibility, ownership and commitment.

- C. *Utility* of evaluations. To have an impact on decision-making, evaluation findings must be perceived as relevant and useful and be presented in a clear and concise way. They should fully reflect the different interests and needs of the many parties involved in humanitarian assistance and/or development co-operation. Ensuring the utility of evaluations is only partly under the control of evaluators. It is also a function of the interest of managers, and member states through their participation on governing bodies, in commissioning, receiving and using evaluations.

The findings of this Review are based on the following:

- Preparatory work included a desk review of relevant documentation and initial interviews at WFP headquarters in Rome.
- Based on these preliminary data collection exercises a preliminary factual report was prepared and circulated to OEDE for verification and comments.
- Field visits were made to WFP Regional Bureaux and Country Offices in Johannesburg, Lilongwe, Jakarta, Bangkok and Vientiane where WFP staff, partners and other stakeholders were interviewed.
- A meta-evaluation of twelve OEDE and decentralised evaluations was conducted. In conjunction with this desktop assessment, interviews were also conducted with evaluation team leaders, evaluation managers, Country Office/Regional Bureau staff and others who were involved with these evaluations.
- This data was compared with a review of ALNAP's assessments of WFP evaluations.
- A web-based survey of the views of WFP staff was conducted which received 87 responses.
- Peer Panel interviews with selected stakeholders were conducted, including: the Executive Board (a meeting held with fourteen members), executive staff, OEDE (director and staff), regional and country directors, division directors, regional monitoring and evaluation focal points, and other WFP staff.
- In total 124 interviews have been carried out.
- The draft assessment report was discussed with WFP in a Review Meeting on October 25, 2007. Based on this discussion the final assessment report has been produced.

The three criteria or dimensions used for the analysis – independence, credibility and utility – relate to both objective and subjective judgements. The UN’s normative framework (the UNEG Standards and Norms, as well as the OECD DAC criteria used in this Review) is based on a mixture of corporate, cultural and perception-based standards. The source information for the Review is therefore drawn from an analysis of WFP’s organisational structure, related financing, corporate managerial practices and the subjective opinions of staff regarding the three dimensions. The latter perceptions relate to WFP’s readiness to utilise evaluation and are hence particularly relevant in an assessment of an organisation’s evaluation function; they have therefore formed a significant component of the source data used by the Peer Panel.

Main findings and conclusions

General

The Panel concludes that the *Independence* of the WFP evaluation function is adequate in comparison to similar organisations; that the *Credibility* of *products* of the WFP evaluation function is uneven, and that the *process* of the function is somewhat more credible but also problematic; and that the criteria of *Utility* of the WFP evaluation function are partially met with regard to contributing to programming but that structures and mechanisms to promote utility are weak in most other respects.

OEDE is a strong unit with committed, well-trained and highly motivated staff. During the past seven years OEDE has invested much efforts on improving WFP’s evaluation function. OEDE is now addressing a number of weaknesses such as the quality of reports, limited attention to strategic evaluation planning, lack of full management engagement and follow-up etc. The Peer Review Panel feels that if these changes are implemented, they will address many of the findings of the Panel.

The evaluation function is of more variable quality at the level of Regional Bureaux and Country Offices. Levels of motivation and of invested resources depend on the interests and priorities of the offices concerned. Given the nature of decentralisation within WFP, OEDE has not been in a position to exert significant influence on the quality of these evaluations. Nonetheless, the findings of both OEDE and Regional Bureaux led evaluations are routinely reflected in the design of follow-up programmes and have some influence on policy formation. All OEDE evaluations are made public and submitted to the Executive Board.

Independence

The Panel concludes that the *Independence* of the WFP evaluation function is quite good in comparison to similar organisations. Though generally satisfactory, the Panel has noted a danger that this independence could be eroded in the future due to a lack of recognition of the importance of evaluation independence among many WFP staff.

The Executive Board and Executive Director fulfil responsibilities regarding the appointment of a professional head of evaluation in an appropriate manner. Evaluation resources are currently safeguarded. OEDE is outside of line management while at the same time sufficiently integrated into WFP leadership structures to facilitate impact. This is seen by the Panel as being largely appropriate.

There are, however, some weaknesses in the independence of the evaluation function. Accountability for the implementation of recommendations is unclear. Some OEDE staff are concerned that their careers may be affected by their evaluation role, which could lead to inappropriate risk averse behaviour in their management of sensitive evaluations. There are also insufficient safeguards to prevent partiality and conflicts of interest amongst external evaluators. The role of Regional Bureaux in both oversight and advisory support to Country Offices has problematic implications for the independence of their role in decentralised evaluation in that their evaluation function involves public critique of programming while they must also ultimately maintain collegial day-to-day relationship with Country Offices.

With some notable exceptions, OEDE usually has full access to information and is free to express its findings. Evaluations take the views of all relevant stakeholders into account, but the evaluation process does not provide for sufficient dialogue with stakeholders outside of WFP.

Credibility

The Peer Panel has assessed *Credibility* in terms of both the evaluation products and the processes through which evaluations are managed. The Panel has assessed WFP evaluations with regard to both formal quality standards and also the perceptions of credibility within WFP at different levels of the organisation. In some cases these perceptions are quite different from the Panel's assessment of quality. Some decentralised and self-evaluations are seen as highly credible within the countries where they have been conducted since they relate to the core concerns of the relevant stakeholders. Several of those evaluations have nonetheless been judged as unsatisfactory by the Panel and indeed by most stakeholders at headquarters due to inherent faults in methods and failure to adhere to basic quality standards. The Panel concludes that the overall credibility of *products* of the WFP evaluation function is uneven, and that the *process* of the function is somewhat more credible but also problematic.

An official evaluation policy should provide the foundation for credible evaluation. Such a policy exists, but it consists of a layered series of documents which detracts from clarity and applicability. Evaluation policy is not sufficiently used to guide practice.

Evaluators and Regional Bureaux have been unclear regarding what is expected in terms of quality due to a lack of specification within OEDE itself and concerns that headquarters' expectations do not take into account resource and time constraints in the field. This first deficiency is in the process of being addressed by drafting of clearer standards and procedures. OEDE also intends to take on a more proactive role in supporting decentralised and self-evaluation, but problems will remain due to the shrinking levels of human and financial resources, especially within the Regional Bureaux.

On the whole the OEDE evaluation function is impartial. The views of all stakeholders are often sought, but there appears to be an uneven emphasis on stakeholders who are more accessible and articulate, with beneficiary views in particular under-represented. There are claims that some hired evaluators have preconceived ideas and that they are influenced by conflicts of interest. A more competitive approach to recruiting external evaluators could counter this problem and increase the chances to find well qualified consultants.

The process of preparing for evaluations, management and advising and supporting teams in the field is handled in a highly professional manner by OEDE. Terms of Reference are generally of good quality, but they are at times too standardised and over-ambitious. This is partially related to weaknesses outside of the evaluation function, as programme goals are often not clear or well structured. There is in most cases a dearth of monitoring data and existing data is often accepted without sufficient scrutiny.

The quality of evaluations is mixed. This may improve with the planned establishment of new quality standards by OEDE in the near future. Particular weaknesses have been noted in methodology and in some crosscutting issues such as human rights, where WFP as whole lacks capacity and guidance. A failure to take into account the cost implications of recommendations, together with factors related to the nature of priority setting in WFP has damaged the credibility of evaluations among some WFP staff.

Partners are engaged in evaluation primarily as either hosts or key informants. The Panel views this as inappropriate as a basis for encouraging two-way accountability and learning. The extent to which stakeholders are consulted in the planning, design, conduct and follow-up of evaluations is patchy and overly concentrated on the implementation stage.

Utility

The Panel concludes that the criteria of *Utility* of the WFP evaluation function are partially met in that it provides a major contribution to the formulation of follow-up programmes, but that structures and mechanisms to promote utility are weak in most other respects. There is insufficient corporate strategic thinking regarding the use of evaluation and where evaluation should fit on the learning-accountability spectrum. Evaluation is rarely used to enhance accountability to partners.

Evaluation is insufficiently integrated into many of the processes by which WFP sets, monitors and analyses policies. WFP rarely uses evaluation as a way of encouraging broader accountabilities in respect to its position within the humanitarian and development communities.

Evaluation is primarily focused on outputs, as opposed to outcomes and impact, which reflects the demands of many stakeholders for practical programming advice. In a narrow sense of contributing to an understanding of how to ‘do things right’, evaluation makes a notable contribution to programme design and management. In a wider perspective of learning about ‘doing the right thing’, performance is not so good. Despite some efforts within evaluations to present evidence that can stimulate greater reflection within WFP over the changing role of food aid, for example, the corporate view of evaluation has tended to focus primarily on its utility for making modest adjustments to existing approaches. A notable exception to this is the interests of the Executive Board which seeks to use evaluation as a major input into their deliberations on the future role and corporate priorities of WFP.

Structures for management response to evaluations at headquarters level are fragmented and require a major overhaul. Management responds to evaluation through the process of programme formulation, but in a broader sense this is weak. New plans in OEDE to tie evaluation closely to logical frameworks may enhance utility through a focus on outcomes, but lack of prevailing understanding and use of logical frameworks within WFP will make this difficult. These problems ultimately stem from the fact that respect for the role of evaluation and a readiness to acknowledge mistakes are reportedly weak at WFP.

The intention to use evaluations is evident as programme evaluations are timed to coincide with formulation of new phases. At decentralised levels there is a close link to utility since there is a direct desire to use evaluation to inform and justify new programmes and phases. Local flexible solutions are often found to use evaluation within a broader learning process. A negative aspect of this may be that the accountability aspect of evaluation, whereby an independent assessment may question the overall relevance of country priorities, may be overshadowed by a concentration on feeding into the design of attractive programming based on documented success.

Evaluation makes an inadequate contribution to overall knowledge building within WFP and virtually none among partners. Access to reports and findings through the website, debriefings, etc., is acceptable, but promotion of the use of evaluation products is not sufficiently proactive. Evaluation is not well integrated into other learning initiatives within WFP. Opportunities are missed to distil and disseminate learning systematically. Evaluation does make a major contribution to building knowledge within the Executive Board, particularly through thematic/strategic evaluations. There is a demand for greater efforts to present syntheses of evaluations conducted. At decentralised levels evaluation is more directly related to programme formulation and learning is somewhat more effective.

Not enough effort has been made to find opportunities to build local capacities. Some deficiencies noted regarding utility relate to the need to rethink stakeholder engagement throughout the evaluation process so as to ensure broader buy-in. The quality of utility is directly related to the ability of WFP to apply findings and recommendations together with partners, but this aspect of the evaluation function is generally very weak.

Recommendations

The Peer Panel's recommendations are intended to suggest ways to improve and amend aspects of WFP's evaluation function where the Peer Panel has identified problems or shortcomings in relation to UNEG's norms and standards or to established evaluation practice.

Evaluation policy

OEDE should develop an evaluation policy that encapsulates the previous evaluation policies and fully meets all UNEG norms and standards for evaluation. It should also be designed as a transparent vehicle for promoting greater communication among internal and external stakeholders regarding the aims and intended utility of evaluations. The role and purpose of and the relationship between self-evaluation, decentralised evaluations, and external evaluation should be studied and clearly articulated not only in the evaluation policy, but in the overarching policy and strategy documents of WFP. The mandate for OEDE should form part of WFP's evaluation policy.

Wider accountability

A key challenge for WFP's accountability is to expand its institutional accountability to include intended beneficiaries via host government and/or NGO Cooperating Partners. OEDE should develop an 'accountability map' of key WFP stakeholders, both internal and external, to help in clarifying roles and responsibilities. Based on this 'accountability map', guidance should be provided both for WFP staff in func-

tional units and partners identified in the map as key stakeholders to help them in fulfilling their accountability responsibilities.

Participation in evaluation

OEDE should look for ways of promoting, and providing incentives for staff to adopt more participatory approaches in evaluations. Engagement with partners at country, regional or global levels is primarily a responsibility of other parts of WFP. Nonetheless, OEDE should play a more proactive role in promoting substantive involvement of relevant stakeholders. The responsibility of the OEDE evaluation manager and evaluation team leader is to make optimal use of participatory systems already in place during the planning, implementation and utilisation of the evaluation. OEDE has a role to play in building capacity of WFP field staff to ensure that WFP field staff are provided with appropriate support and guidelines to facilitate participatory approaches during evaluation processes.

Management response to evaluations

The lines of responsibility for management response are currently blurred. WFP should, both in principle and in practice, establish a clear division of responsibility regarding management response between the evaluation function and the organisation's line management. After an evaluation has been submitted to the Executive Director, OEDE should not be involved with drafting or compilation of responses from different parts of the organisation. The management response mechanism should include rules about the timeframe for the response and procedures for follow-up of the management response as well as for reporting to the Executive Board and informing the OEDE about the results of the follow-up.

A similar system for management response should be used for decentralised evaluations. The same kind of division of responsibilities cannot be established when a country director both commissions an evaluation and decides on management response, for example. However, when an evaluation is carried out by external evaluators a country director can still independently formulate a management response and report to the Regional Bureau director or the Executive Director. Follow-up should also be the Country Office's responsibility with reporting on the results upwards within WFP. Management response and follow-up mechanism should be transparent with relevant documents easily accessible for WFP and partners and routinely posted in electronic form.

Quality of evaluation

The increasing emphasis on improving the quality, rigour and harmonisation of OEDE's work, as well as the focus on systematic processes, quality checks and tools such as stakeholder maps and evaluation matrices are highly encouraged. The capacity of OEDE staff should be maintained over time to stimulate interest in the evaluation field and encourage professionalism. We recommend that ample time should be allocated and incentives should be provided for staff to keep up with new developments in the field of evaluation.

Mechanisms should be found to improve the quality, credibility and ownership of evaluation recommendations. Such mechanisms may include developing recommendations in dialogue with primary stakeholders, and/or leaving recommendations up to those responsible for decisions and action in WFP, based upon engagement by primary stakeholders around the findings and conclusions of the evaluation report.

Learning and accountability

The tension and complementarities between evaluation for accountability and for learning seem not to be acknowledged everywhere in WFP. This requires more attention. Ideally the learning element should be linked to a larger organisational knowledge management strategy.

OEDE should continue recent efforts to systematically harvest lessons from existing evaluations as well as external fora such as ALNAP, the IASC and relevant partners. Innovative methods for extracting and sharing of evaluation lessons should be investigated, building on the experiences of other organisations with extensive experience in this field.

Results-based management

WFP should give high priority to address the fragmentation among its various results-focused data collection, reporting and analysis tools. A thorough review of existing field monitoring systems and applications is vital to ensure that evaluations as well as the corporate monitoring system have access to more reliable, relevant and comparable data. Ways should be developed and maintained to ensure that all interventions are linked to proper monitoring mechanisms, both at local and corporate levels. OEDE should make an evaluation of WFP's approach to results-based management a high priority for its future strategic evaluations, building on the recent work done by internal audit on results-based reporting.

Selection of evaluation teams

It is recommended that OEDE develops a more transparent, rigorous and competitive approach to the selection of team leaders. If possible, team leaders should be identified early on and be involved in the identification and selection of the rest of the team. All evaluation teams should include at least one evaluation specialist, preferably the team leader, who has sufficient knowledge about and experience from current evaluation approaches and methods.

Staffing and budgeting of OEDE

The Panel considers the current mix of internal WFP career staff and externally recruited professional evaluators suitable. However, given the technical nature and professional skills profile of these positions, the Panel recommends WFP to allow OEDE to select internal staff based on a professional recruitment process rather than through the standard reassignment exercise. WFP should continue to (a) allow external recruitment of evaluation specialists; (b) base OEDE's staff profile on the profile of evaluators developed by UNEG; and (c) consider how to ensure an appropriate career path for evaluation specialists within WFP and within the UN system.

To address concerns that only a small portion of the overall evaluation budget is within the direct control of OEDE, WFP's senior management should devise ways to safeguard the funding allocated to evaluations for the next biennium. The use of Direct Support Cost of projects appears a budgetary necessity for the time being but it is not an ideal situation. It is critical that the Executive Director and senior management ensure the full use of these funds for evaluation by holding managers accountable for following implementation plans. Furthermore, management should consider 'earmarking' strategic and sensitive decentralised evaluations for OEDE-management to thus reduce the risk that offices seek to bypass OEDE and to ensure full independence where most needed. The establishment of a centrally managed fund for both OEDE evaluations and decentralised evaluations should be investigated.

1. Introduction

1.1 Background and purpose of the Peer Review

On 20 December 2006, WFP formally requested the Chair of the DAC/UNEG Task Force for Professional Peer Review to initiate a Professional Peer Review of the Evaluation Function at WFP. The purpose of the Review was to provide the WFP leadership, the member countries through the Executive Board and OEDE with an independent assessment that could inform ongoing organisational planning processes and decisions on the function's positioning and work. The Review was to be the third in a process launched to help improve the evaluation work of UN and DAC members. It was therefore to be conducted and used by members in line with similar reviews completed in UNDP and UNICEF in 2005 and 2006 respectively.

The Office of Evaluation at WFP describes the recent development of its evaluation function and the decision to initiate a Peer Review as follows:

“WFP has had an evaluation function since 1965. The Office of Evaluation (OEDE), and its predecessors, has undergone a number of changes since. Between 2000 and 2005, the Executive Board of WFP received four papers on the evaluation function of WFP¹. The main concerns of the Executive Board members were (i) location, (ii) reporting lines, (iii) budget, and (iv) interface with the Board.

During the Board's 2006 annual session, the Executive Director of WFP agreed to professionalizing OEDE by (a) appointing a director with a proven track record in

¹ WFP's Principles and Methods of Monitoring and Evaluation, presented for consideration at WFP/EB.A/2000/4-C; A Policy for Results-Oriented Monitoring and Evaluation in the World Food Programme, presented for approval at WFP/EB.A/2002/5-C; WFP's Evaluation Policy, presented for approval at WFP/EB.3/2003/4-C; and Report on the Management of Evaluation, presented for consideration at WFP/EB.A/2005/5-E. The four documents can be accessed at www.wfp.org/operations/evaluation.

evaluation (rather than filling the position through an internal appointment), (b) upgrading the position and making the director of evaluation part of the executive staff, (c) changing the reporting lines, i.e. the Director reports now directly to the Executive Director.

At the same time, WFP informed the member countries of WFP about the intention to volunteer for a professional peer review of the evaluation function. As a result, the Executive Board members agreed to hold further queries and discussions about the independence of OEDE until the results of the Professional Peer Review were presented to the Executive Board.”

1.2 The Review process

The Department for Evaluation and Internal Audit of Sida agreed at the end of January 2007 to lead the Review. Over the next few months the process was planned in conjunction with OEDE, the Peer Panel established and consultants recruited as advisors. The Panel had its first working meeting in Rome in April, followed by several more until the conclusion of its work in November 2007. It reported from time to time on progress to OEDE and the DAC/UNEG Task Force.

Based on a first draft report and in line with the interactive nature of the Peer Review process, discussions to test preliminary and potentially controversial findings took place on 25 October with OEDE staff as well as management representatives. Based on this discussion and comments received by the Peer Panel, the draft was edited for factual accuracy and submitted on November 5th to OEDE which in turn is to provide it to the Executive Director of WFP and disseminate it within WFP. The final assessment report is to be provided for information to the DAC/UNEG Task Force and will be publicly available. The report is to be presented by OEDE to the WFP Executive Board’s 2008 first regular session in February 2008.

The Panel received excellent and proactive support both at headquarters and in the field from OEDE and from other WFP stakeholders and evaluation team leaders. Full access was given to all requested documentation, as well as assistance in identifying and arranging contacts with stakeholders. The Panel was able to act in an independent manner without any interference from WFP staff. It found a remarkable level of openness within the organisation.

1.3 The Peer Panel members and advisors

The Peer Panel consisted of volunteers prepared to give their time to the Review. It was constituted to reflect the following:

- Relevant professional experience. WFP has a dual role in providing humanitarian and development assistance. It was therefore impor-

tant that Panel members brought together professional experience of both types of work.

- Independence. In order to avoid any conflict of interest or partiality, none of the Panel members was to have a relationship with WFP that could influence the Panel's position and deliberations.
- Representative membership. The experience and interests of donors, executing organisations and partner countries were to be represented on the Panel.

The following members² agreed to the task:

Jock Baker,	Programme Quality and Accountability Coordinator, CARE International
Stefan Dahlgren,	Senior Evaluation Officer, Sida (Chair)
Susanne Frueh,	Chief, External Relations, Peace building Support Office, United Nations, former Chief of Evaluation, OCHA
Ted Kliest,	Senior Evaluation Officer, Netherlands Ministry of Foreign Affairs
Zenda Ofir,	International evaluation specialist, Evalnet, South Africa

The Panel was assisted by two advisors, Ian Christoplos and Peta Sandison, both with broad experience in humanitarian and development evaluation, who were responsible for the collection of (primarily factual) data and information in support of the Review.

1.4 The focus of the Review

The Professional Peer Review takes the central evaluation function, i.e. OEDE, as its starting point, but recognises that it does not operate in isolation. Aspects of the decentralised evaluation activities and the engagement of the organisation as a whole in evaluation had to be considered. This meant that the work of other relevant units in WFP as well as the attitudes and perceptions that frame how WFP staff at different levels view and use evaluation had to be understood.

The Review had to be in line with the *Framework for Professional Peer Reviews of Evaluation Functions in Multilateral Organisations*, which highlights three carefully selected core criteria that need to be satisfied for evaluation functions and products to be considered of high quality:

- D. *Independence* of evaluations and the evaluation system(s). The evaluation process should be impartial and independent in its function from the process concerned with the policy making, the delivery, and the management of assistance. Independence of the evaluation function is a pre-condition for credibility, validity and usefulness. At the same time, the review recognizes that the appropriate guarantees of the necessary independence WFP is defined according to the

² A short description of the professional background for each Panel member and for the advisors is in Appendix 4

nature of its work, governance, decision-making arrangements, and other factors. Moreover, like most organisations WFP's aim is to encourage the active application and use of evaluations at all levels of management, meaning that systemic measures for ensuring the necessary objectivity and impartiality of this work should receive due attention.

- E. *Credibility* of evaluations. The credibility of evaluation depends on the expertise and independence of the evaluation managers and the evaluators and the degree of transparency of the evaluation process. Credibility requires that evaluations should report successes as well as failures. Partner countries should, as a rule, fully participate in evaluation in order to promote credibility and commitment. Whether and how the organisation's approach to evaluation fosters partnership and helps build ownership and capacity in developing countries merits attention as a major theme.
- F. *Utility* of evaluations. To have an impact on decision-making, evaluation findings must be perceived as relevant and useful and be presented in a clear and concise way. They should reflect the different interests and needs of the many parties involved in development cooperation. Importantly, each evaluation should bear in mind that ensuring the utility of evaluations is only partly under the control of evaluators. It is also critically a function of the interest of managers, and member countries through their participation on governing bodies, in commissioning, receiving and using evaluations.

The Review was somewhat complicated by the ongoing organisational restructuring which made it at times difficult to identify and locate relevant functions and initiatives affecting the evaluation function. In spite of this, the three core criteria as well as the WFP mandate, objectives and related organisational functions provided the Review framework. It was further defined through a set of questions, based on the UNEG norms and standards that the Review sought to answer³. The key issues to be addressed by these questions are summarized as follows:

1. Structural aspects of how the evaluation function operates in WFP, in particular whether the current functional arrangements are effective in ensuring independence and that evaluation can contribute to both learning and accountability within WFP.
2. The WFP evaluation policy as well as other policies and strategies with a bearing on OEDE and its work – in particular the extent to which the evaluation policy and/or strategy conforms with international standards, and the relevance of other policies to the functioning of OEDE (e.g. on results-based management, strategic planning, budgeting, decentralised evaluations, etc.).

³ These questions will be made available to UNEG and DAC's Evaluation Network

3. The organisational relationships of OEDE at the governance level, with reference to the Executive Board and Bureau, Executive Director, the Core Management Team, and the Executive Staff; and the roles and responsibilities of OEDE in relation to other HQ departments (e.g. Policy, Operations).
4. The decentralised approaches to evaluation, including the roles and responsibilities of OEDE vis-à-vis the Regional Bureaux and the Country Offices.
5. Relationships with, and responsibilities vis-à-vis WFP's cooperating partners, including the nature of stakeholder engagement before, during and after the evaluation process.
6. The quality of the evaluations undertaken and commissioned by OEDE and to some extent by Regional Bureaux/Country Offices. This includes the conduct of the actual evaluation, the quality of the evaluation reports, the independence of evaluation teams and team leaders (consultants), the ways in which OEDE enables them to produce credible reports including the ways stakeholders are facilitated to comment on draft reports (e.g., when do comments become an infringement on independence and when are they warranted to ensure standards of evaluation reports).
7. The use of evaluation results and follow-up. Important aspects: the ways in which evaluation results are disseminated and lessons used both within WFP and by others (donors, cooperating partners, etc.); the responsibility for the follow-up of recommendations with management; and how follow-up is undertaken and monitored.

1.5 Methodology

While the Review framework and questions guided the aspects to be investigated, the methodology was designed using a mixed methods approach to (i) allow new and important issues to emerge, with additional data collection done as the Panel's understanding deepened; and (ii) enable triangulation (cross-checking to verify information) based on data collection methods as well as the sources of information. Due to the nature of the Review the methods employed were primarily qualitative. An inductive approach – seeking patterns rather than testing hypotheses – guided the gathering and analysis of information.

The three criteria or dimensions used for the analysis – independence, credibility and utility – relate to both objective and subjective judgements. The UN's normative framework (the UNEG Standards and Norms, as well as the OECD DAC criteria used in this Review) is based on a mixture of corporate, cultural and perception-based standards. The source information for the Review is therefore drawn from an analysis of WFP's organisational structure, related financing, corporate managerial practices and the opinions of staff regarding the three di-

mensions. The latter's perceptions relate to WFP's readiness to utilise evaluation and are hence particularly relevant in an assessment of an organisation's evaluation function; they have therefore formed a significant component of the Review's source data.

A total of 124 persons were interviewed. Interviewees were not randomly selected but chosen to represent views from key parts of the organisation and, more importantly, to express opinions on and provide experience from evaluation that clearly reflect the use of evaluations and how they may influence decisions. The locations of the field visits and meta-evaluations guided the selection of many of the interviewees. Others were selected by the Panel, in consultation with OEDE, as representatives of the organisational units with which OEDE interacts or which are key stakeholders in the evaluation function. The Panel further selected external stakeholders based on the same criteria. Together they provide, in the Panel's opinion, a rich reflection of the views on evaluation within WFP.

Key steps of the Review:

1. *Preparation of the Review Approach and Work Plan*

This was done in dialogue between the Peer Panel, the Review advisors and the OEDE. The Peer Panel conducted a first desk study, followed by the development of the Review framework and questions during and after the first meeting of the Panel. A work plan was established during the first meeting in Rome, which also served to familiarize the Panel and advisors with the evaluation function of WFP.

2. *Collecting factual information*

A more extensive desk review by the advisors followed. This was guided by the Review questions to gather factual information and identify pertinent issues for further investigation. The desk study was supplemented by semi-structured interviews by the advisors with key persons in Rome. This resulted in a preliminary draft report with factual information provided to the Peer Panel and to OEDE for verification and comment.

3. *Wider consultation within WFP*

Further data collection from WFP staff was carried out through a web-based survey sent to 482 WFP staff.

4. *Assessing the quality of WFP evaluations*

A meta-evaluation of twelve OEDE and decentralised evaluations was conducted by the advisors, based on a modified version of the ALNAP Quality Proforma. They were selected to (i) include a mix of Country Office and OEDE managed evaluations, (ii) ensure a geographical/regional spread, as well as a mixture of (iii) programme and (iv) evaluation type. The results of another 17 meta-evaluations by independent ALNAP assessors between 2000–2004 were also considered. Interviews

to further elucidate issues of quality were conducted (mostly by telephone) with evaluation team leaders and evaluation managers as well as Regional Bureaux/Country Office staff.

5. *Deepening understanding of perspectives from the field*

Visits were made to WFP Regional Bureau and Country Offices in Johannesburg, Lilongwe, Jakarta, Bangkok and Vientiane where WFP staff, partners and other stakeholders were interviewed. These locations were selected based on the convergence of a number of factors: (i) geographical coverage; (ii) inclusion of meta-evaluation countries and regions; (iii) date of the evaluation between 2005–2007, to focus on the most recent situation as well as facilitate staff recall and availability; (iv) type of programme (PRRO, EMOP, CP and Regional); (v) the type of evaluation or review (Mid-Term Evaluation, Real Time Evaluation RTE, After Action Review AAR); and (vi) timing of the visits (availability of the advisors and relevant Country Offices to host the visits).

6. *Testing and deepening understanding of facts and perspectives:*

Peer Panel interviews in Rome

The Peer Panel interviewed in person selected stakeholders at WFP headquarters, as well as, by telephone, key persons in the field. These interviews focused on further investigating key issues raised during the earlier steps of data collection and analysis. A group discussion was held with fourteen members of the Executive Board, present after an open invitation to attend the engagement. Individual semi-structured interviews were conducted with Executive members, Division Directors, several Regional and Country Directors and staff, as well as OEDE.

7. *Verification of preliminary findings*

A preliminary assessment by the Peer Panel was made based on all the evidence gathered through the different methods and captured in a draft Review report which was then submitted to OEDE for their comments. The main Review meetings took place on 25 October, where the Panel interacted with OEDE and several key WFP officials on the draft report content. Opportunity was thus given to test and refute the findings.

8. *Finalisation of the Professional Peer Review Report*

The Review meetings brought to light additional aspects for crosschecking and for final consideration. This was done and the final Professional Peer Review report submitted to OEDE on November 5th.

1.6 Limitations and challenges of the Review

1. It was difficult to find staff and stakeholders at field level who had been directly involved in the evaluations under review. Many of the sources had fairly scattered and limited experiences of evaluations. The rotation of personnel in the organisation and limited recall of

specific evaluations made it more difficult than expected to gather stakeholder perspectives.

2. It was also more difficult than expected to conduct productive interviews with key external stakeholders, in particular government and NGO cooperating partners. They had been only marginally involved in evaluation processes, or not at all. A key challenge for WFP's accountability emerged during this process, in that institutional accountability tends to be regarded as primarily "upwards" towards the Executive Board and major donors. Its sense of accountability towards intended beneficiaries is less strong and dependent on the interface between the host government and/or NGO cooperating partners.
3. Neither the Peer Review's nor ALNAP's meta-evaluation of evaluation reports were weighted with respect to particular aspects of evaluation quality. In retrospect the Panel concluded that some factors could have been given greater prominence as preconditions for overall credibility. For example, a report's limited presentation of its evidence-base or its methodology would have been given additional weight, since such weaknesses could mean the evaluation is fundamentally flawed and hence judged as unsatisfactory overall. It is recognised that in some cases, a more extensive description of methodology is included in the ToR and the Annexes; however this is not standard practice. Under these circumstances it was difficult to determine the quality of the evaluation. This significantly limited the credibility and usefulness of the meta-evaluations.
4. The web survey had a response rate of only 18%, which, given the sample size, was not sufficient to draw firm conclusions. However its results were crosschecked during interviews, and these largely concurred.
5. The time available for the Peer Panel to do a thorough crosscheck and systematic analysis of all the collected data and information, especially of new issues that emerged during the last round of interviews was limited. This meant that triangulation was lacking in some instances and made the final Review meetings a very important part of the validation process. The draft report was adjusted to reflect additional insights gained through the verification process.

In spite of these constraints and challenges, a fairly consistent picture emerged on most of the key issues. Although not all aspects were covered by this Professional Peer Review, the Panel is confident that the report can serve as a credible input and stimulus for WFP as it moves forward to improve and embed the evaluation function as a critical component in its ongoing search for excellence in fulfilling its mandate.

1.7 Organisation of the report

After the introductory chapter on background and approach and a chapter briefly describing the evaluation function in WFP, the report discusses in turn each of the criteria or analysis dimensions in three separate chapters. Different facets, pertinent to the dimensions of independence, credibility and utility, are described under separate headings and examined regarding their importance.

Each of the three main chapters ends with a brief, overall conclusion and a table with the relevant UNEG norms and specific findings related to that norm.

The report ends with a chapter on recommendations, organised along the main issues the Panel identified during the analysis. When applicable and for practical purposes the recommendations are directed to either WFP or to OEDE but they are of course all, in principle, addressed to the organisation as a whole.

2. Evaluation at WFP

2.1 The current evaluation system in WFP

Evaluation is a long-standing function in WFP and is carried out by the Office of Evaluation (OEDE) and at the operational level by Country Offices and Regional Bureaux (decentralised evaluations).

The evaluation system in WFP is best described in the 2002 M&E Guidelines which define three different types of evaluations:

- OEDE-managed evaluations
- Country Office and Regional Bureau-managed evaluations (decentralised evaluations)
- Country Office self-evaluations

The guidelines for Country Office self-evaluation include a type of After Action Review (AAR) or facilitated annual review of project performance. The AAR concept has been further developed by Policy into a stand-alone function.

OEDE output is fairly steady at 10–12 evaluations per year (similar to the output in 2001). Following the decline in development-oriented food assistance OEDE has, since 2002, increasingly focused on ‘large-scale humanitarian assistance interventions’ and ‘strategic’ evaluations (such as thematic evaluations). This shift has meant that the responsibility for assessing the quality of specific programmes has been largely decentralised to the regional and country level.

2.1.1 A strong central evaluation function

At present OEDE consists of a Director and seven professional staff. This figure has remained fairly constant over the past decade (in 1999 OEDE consisted of a Director and six professional staff). Three of the professional staff positions are at senior level (P-5), the remainder at P-4 level. OEDE staff and non-staff costs are all funded from the PSA

budget which funds all core regular staffing and support costs of the organisation. OEDE sometimes seeks additional earmarked funding for specific activities from donors. For example one of the team members on the Southern Africa PRRO evaluation was funded by a donor. The M&E Guidelines were funded by DFID under the first Institutional Strategy Paper (1999) which, for a short period, also enabled the development of an M&E roster and the proffered deployment of M&E advisors at no cost to offices. In addition, over the last six years, various donors have funded one junior professional position in OEDE.

2.1.2 A weaker decentralised evaluation function

The evaluation system foresees that Country Offices and Regional Bureaux undertake decentralised operation-specific evaluations. Each Regional Bureau has an M&E Focal Point, but this is not a dedicated post. The quality and intensity of engagement between the field and OEDE is highly variable depending on the individual Focal Point, their relationship with the relevant OEDE Evaluation Officer and the level of competing pressures and priorities. These factors, OEDE's overall capacity and the profile of the Focal Point in the Regional Bureaux largely determine the ability of OEDE to support decentralised evaluation.

OEDE provides assistance for decentralised evaluations on a demand-driven basis. OEDE responds to requests for comments on the ToR, refers to guidelines and shares names of potential consultants. The support provided is generally more opportunistic, based on contacts that arise with the field office in question, rather than as a result of systematic contacts and follow-up.

Despite the calls for a shift to more decentralised evaluations, there is a significant gap between the numbers of decentralised evaluations contained in regional work plans and those actually carried out. The reason given is usually lack of budgetary resources, but this could also be interpreted as an indicator of the priority accorded to evaluation. There was previously a discussion of creating a fund for decentralised evaluations but it was decided this would undermine the principle of regional responsibility. This debate is currently being revisited. The direct support costs (DSC) of an operation include an M&E budget line, but the interpretation of this line item is quite loose. Items funded include equipment (such as vehicles) and entire sub-offices, since any cost that relates to the task of monitoring the delivery of food aid can be charged to this line.

2.1.3 Corporate support structure

In recent years two groupings have been added at corporate level. In an effort to address mostly project design weaknesses, WFP instituted the interdivisional Programme Quality Assurance team (PQA) made up of one or more representatives from ODA, PDP, OEDE, OEDR (now dis-

solved) and the Regional Bureau (Programme Adviser). It is chaired on a rotating basis by the Head of OEDE, Director of Policy Development and Associate Head of Operations. The PQA provides a forum for interdivisional dialogue on key programme quality issues and products. The PQA has, *inter alia*, discussed proposals to replace decentralised evaluation with AAR and reform of the Programme Review Committee ToR and process.

The 2005 Report on the Management of Evaluation foresees the establishment of a consultative Internal Evaluation Committee (IEC) composed of executive staff and chaired by the Director of Policy. It is intended to input into OEDE's preparation of its work programme and discuss systemic issues arising from evaluation and how to address them. A core question has been how to more effectively deal with decentralisation and the increasing responsibility of field offices for evaluation and utilisation of evaluation results.

2.1.4 Use and application of evaluation findings and recommendations

The current system places strong emphasis on the use of evaluations. Evaluations are therefore typically timed to feed into the assessment and design of follow-up programmes (primarily PRROs and EMOPs) and into new policies (thematic evaluations). OEDE has an opportunity to ensure that evaluation findings are reflected in new design when it participates in the Programme Review Committee (PRC). The PRC is a specific grouping organised around each new programme. Its role is to provide policy and normative advice on the design of all operational programmes and to ensure that quality control of and consistency between the different programmes is maintained. The PRC constitutes one of the main central opportunities for promoting learning and accountability through applying the findings and recommendations of evaluations. Members post comments on an internal comments intranet site and then meet to review. OEDE provides input, but like other contributors does not have authority over whether new programmes are accepted.

The main mechanism for follow-up is a management response to a matrix of the recommendations. Under the current Evaluation Policy, OEDE does not have the responsibility for ensuring or tracking compliance. Its mandate is to obtain the management response and submit it to the Executive Board. In general, responsibility for follow-up lies with the Country and Regional Directors following acceptance by the Executive Board. The Executive Board occasionally requests further clarifications and, more recently, subsequent management responses.

OEDE is currently designing a new recommendation tracking system. The software is expected to be developed in October 2007 to be operational at the latest by the end of December 2007. A consultant has been recruited for this task.

2.1.5 Evaluation reports

As a matter of policy, all OEDE-commissioned evaluation reports are made public by posting them on the website. Since 2004, OEDE has issued an Annual Evaluation report which includes lessons identified as well as a review of compliance, but it is acknowledged by OEDE that the dataset is weak and does not allow full compliance monitoring.

All OEDE-managed evaluation reports are presented to the Executive Board. A summary report of ten pages preceded by a one-page abstract is presented. Full ‘technical’ reports are not shared in hard copy with the Board prior to the discussion; they are posted on WFP’s website. If a new operation or policy update (e.g., on gender or HIV/AIDS) is being presented, OEDE presents the evaluation in the same session. If there is no regional connection (e.g., thematic evaluations), then OEDE may present the report in a separate session.

Decentralised evaluation reports are not presented to the Executive Board; they are however analysed for OEDE’s Annual Report.

2.1.6 Types of evaluations

WFP evaluates a number of different programme types (e.g. country programmes, stand-alone development projects, PRROs, EMOPs) using a number of evaluation approaches (e.g. thematic, RTE, self-evaluations). The selection of evaluations presented in OEDE’s biennial work plan follows a consultation process with the Regional Bureaux and Country Offices. Ex-post and final evaluations are rare as are impact evaluations generally. Most evaluations are undertaken several months before the current phase is over.

The 2003 Evaluation Policy stipulates that:

- any operation longer than 12 months should be evaluated;
- OEDE would continue to ensure an independent evaluation service to WFP’s Executive Board, focusing on evaluating corporate programme and policy issues, large operations, and first-generation country programmes at their mid-point;
- OEDE would identify and disseminate lessons and knowledge gained through evaluations to support improved programming and organisational learning; and
- OEDE would provide guidance and support to Regional Bureaux and Country Offices to permit effective implementation of the policy.

The term ‘evaluation’ as per the 2003 policy, includes self-evaluations by Country Offices, evaluations managed by the Regional Bureau or Country Office and OEDE-managed evaluations.

2.2 Approach to evaluation

2.2.1 Conduct of evaluations

Evaluation teams are recruited through networks, from individual consultants known to OEDE and from recommendations (followed through with curriculum vitae). According to OEDE, Team Leaders are generally interviewed by the OEDE evaluation manager. An open, competitive process of recruitment such as a call for expressions of interest is not practiced because of a stated preference to hire consultants individually rather than from or through a company. OEDE staff explained that such contracts would be more expensive than OEDE's budget allows and more time-consuming to manage. In addition OEDE evaluation managers report that directly choosing individual team members results in a greater measure of quality control.

The process for conducting each OEDE evaluation is not formalised. Some guidance is provided by the M&E Guidelines and OEDE is currently working on process maps that will detail and further formalise processes. In general, the *de facto* practice is that, following the selection of the evaluation, the terms of reference are drafted by the responsible OEDE evaluation manager. They are then widely shared with key stakeholders such as Operations Department and Policy in headquarters and the Regional Bureau and Country Office concerned. ToR are usually shared with the Team Leader of the evaluation for his or her input, (more rarely with the team, who may not have been recruited at that point). Comments are then incorporated, largely at the discretion of the OEDE evaluation manager. The Country Offices are responsible for sharing the ToR with partners; the Peer Review sample indicates that this is not routine but does occur, for example in the Bhutan, Niger and Haiti evaluations. It is also common for the ToR to be finalised during a pre-mission field trip if one is taking place, or upon the evaluation team's arrival in the Country Office or Regional Bureau should they have been unable to meet prior to the evaluation's starting.

The practice of requesting an inception report from the evaluation team is increasing, but is not yet formal practice. Debriefing is typically provided by the evaluation team prior to departure from the country and accompanied by an Aide Mémoire. There may be separate internal and external debriefs. The team may also debrief in the Regional Bureau and will debrief again in Rome. Other Regional Bureaux may participate in the teleconference debriefs and similarly, other departments (from Rome) may participate in regional debriefings. The draft evaluation report may go through several iterations. It is distributed to the stakeholders in the country and regional programmes and key stakeholders in headquarters.

2.2.2 Linkages to other HQ functions

There is collaboration between OEDE and the Policy, Strategy and Program Division (PDP) particularly regarding thematic evaluations and participation in the IEC.

Respective Programmes of Work are shared between OEDE and the Division of Oversight Services (OSD). OEDE evaluation managers consult audit reports when preparing an evaluation. The Director, Oversight Services Division and Inspector General (ODS) is a member of the Internal Evaluation Committee. Other opportunities for relations with other HQ units are: the PRC; Programme Quality Assurance Group, participation in workshops or discussions on subjects related to focal point responsibilities; and consultations at various stages of an evaluation (e.g., concept note, ToR).

OEDE appears to currently enjoy good relationships with other WFP staff, though there have apparently been strained relationships with operational staff in the past, particularly at headquarters. Tensions in the field have also been apparent between evaluation team leaders and Regional and Country Directors, but have had somewhat less ramifications on the latter's relations with OEDE.

3. Independence

3.1 Role of the Executive Board and Executive Director

The Panel noted the general satisfaction within WFP (including OEDE itself) regarding current reporting lines. The OEDE Director now reports directly to the Executive Director, and submits all its evaluation reports directly to the Executive Board. Continued commitment and ownership by senior WFP management is considered essential by the Panel for OEDE to retain its relevance within the organisation and the Panel believes that this outweighs concerns expressed in some quarters that OEDE may be subject to the interests, priorities and concerns of the Executive Director.

Some Executive Board members would prefer a model (used by IFAD) whereby the evaluation office reports to the Board directly, but there does not seem to be agreement within the Executive Board whether there is a need to change the current set-up. One member expressed the desire for the Executive Board to be consulted on the choice of the OEDE Director (which was not done in the current appointment) and on any eventual termination. Executive Board members also indicated that they appreciated the model used in IFAD whereby the President of IFAD submits a report to its Board on the follow-up to evaluation recommendations and management actions.

The OEDE is generally perceived as, in effect, being ultimately more accountable to the Executive Board, with the reporting channel to the Executive Director more of a formality. OEDE is working to maintain a balance between maintaining its independence from management by fostering a strong relation to the Executive Board in exercising accountability functions, with increasing attention to stronger links to the Executive Director and management to enhance the impact of evaluation on policies and programming. The proposed Strategy and Work plan strongly emphasise learning and informing programming. One aspect of this emphasis is that accountability is effectively downplayed

(including the relationship to the Executive Board), which may suggest some cause for concern regarding the learning-accountability balance.

The Panel concludes that the current direct reporting line between the Director of Evaluation to the Executive Director is adequate. It does not seem to negatively impact the “independence” of the evaluation office. Noting that some Board members seem to prefer a more direct reporting line to the Board, the Panel is of the opinion that increased Board oversight could be established without changing existing reporting lines through the establishment of a dedicated evaluation sub-committee of the Board. On balance the Panel concludes that OEDE has a greater chance of affecting change and learning from within the organisation and therefore recommends maintaining current reporting lines.

3.2 Independence in programming, carrying out and reporting of evaluations

The Panel observes that OEDE has the appropriate authority to propose its own programme of work, which is ultimately approved by the Executive Director. The emphasis has been on major humanitarian operations and other larger programmes which are either due or overdue for evaluation. There is strong interest from the Executive Board in strategic (e.g. thematic) evaluation and this, together with dialogue with Policy, contributes to OEDE’s decisions regarding programming such evaluations.

In the view of the Panel the emphasis on major operations and themes risks that some small but strategically enlightening operations are not evaluated. OEDE intends to address this by changing the current method of programming evaluations by creating a more strategic and representative procedure for the programming of evaluations. The Panel suggests that it may be more useful to identify systemic or otherwise important issues from the Annual Report or from an Executive Board request and choose priorities derived from that. The current policy directive of evaluating any operation over 12 months is not practical (there are more than 125 of these) nor strategic. Another key aspect of OEDE’s evaluation portfolio concerns evaluations of all new country programmes at mid-point. Country programme evaluations by definition only look at the development portfolio but do not include any PRROs and EMOPs. To ensure a more holistic view of WFP’s activities in a given country and in order to address the connectivity between these, OEDE will introduce ‘country-level’ evaluations which will cover the full ‘portfolio’ of all WFP operations as well as any other activities, such as Vulnerability Analysis Mapping (VAM) during a given time-frame. The Panel supports such a move.

It is unusual for field offices to refuse or contest proposed evaluations. This usually only occurs if there are too many other activities or pressures at a given time. There may be less interest in participating in thematic evaluations since the Country Office or Regional Bureau does not have the benefit of having their own programme evaluated. The financial and human resource burden of decentralised evaluation for Regional Bureau may result in a degree of preference for OEDE-led evaluation.

OEDE staff act as evaluation managers and, in many cases, as quasi team members with a resource role designed to enhance access to information and knowledge of WFP's systems. Evaluation Team Leaders (consultants) were generally extremely positive about this participation and had not experienced it as interfering with their independence. The Panel notes however that the exact role of the OEDE staff member in the evaluation process (regarding guidance, quality management or participation in data collection) and their contribution in terms of authorship of the report is not always clear and needs to be clarified.

As discussed further in relation to credibility and utility, the participation of key non-WFP stakeholders in the evaluation process is often limited. The Panel feels that this can weaken the impartiality of the findings, and hence undermine independence.

In general the Panel found that independence is not a major concern among many WFP staff, perhaps reflecting the 'get on with the work' organisational culture. Independence may thus be threatened by the lack of attention given to evaluation's role in driving accountability within the organisation. According to interviews, evaluation has a low profile and is not seen by many as having a unique role or value in the organisation. This lack of respect may effectively erode norms that would safeguard independence.

The reporting of evaluations to the Executive Board is seen by many of those interviewed as ensuring independence. However some staff from Operations feel that the Executive Board gives OEDE evaluations a level of credibility that is not justified by the quality of the evaluations themselves or the appropriateness of all the evaluation recommendations.

There is less emphasis on independence for the decentralised evaluation function. The Regional Bureau staff see their role in conducting evaluation as part of their oversight role *vis-à-vis* Country Offices. Regional Bureau staff at times lead or participate in evaluations. While they are not directly related to the operation, it is not clear to what extent Regional Bureau staff who participated in the operation's design are systematically excluded from participation on the evaluation.

There is generally less concern about independence at the field level, which in part reflects the Regional Bureaux' combined role of oversight and support. The staff must maintain the independence and objectivity needed to assess the work of the Country Office while also fostering a collegial relationship. A second, and perhaps more worrying factor, is that of relatively junior Regional Bureau staff being put in a direct oversight function vis-à-vis senior Country Office Directors who are at times not averse to using their position to pressure Regional Bureau staff to change their findings.

The Panel feels that more guidance to and training for Regional Bureaux and Country Offices is needed to ensure that any potential conflict of interest is avoided and that undue pressure on external evaluators is minimized.

3.3 Independence and impartiality of evaluation managers and evaluators

The power relations within WFP between Country Office and Regional Bureau Directors at P-5, D-1 and D-2 levels and forceful senior Operations management has meant that it is not uncommon for pressure to be put on evaluators more generally. The previous OEDE Director appears to have taken a strong stance to protect external evaluators and their OEDE evaluation managers from such pressures but his relatively weak position within the organisation reportedly limited his ability to do so. Evaluation teams report varying levels of support from OEDE evaluation managers in defending their independence. One team leader complained of evaluation managers taking a negotiative stance, whereas others were generally pleased with OEDE support. With regard to decentralised evaluation, no such support seems to be possible from HQ given the limited role played by OEDE.

Several interviewees internal and external to OEDE acknowledged that OEDE posts are highly stressful regarding relations between staff and other members of the organisation. Interviews showed a range of perceptions regarding how these pressures impact on independence. Some felt that a posting in OEDE was a 'bad career move' and might affect the individual's future career, whereas others thought that these dangers were manageable. A potential problem is noted with WFP's current practice related to staff selection for vacant posts and promotion. The recommendations of the Professional Promotions Panel are reviewed and communicated to the Executive Director by an intermediate senior management panel. Hypothetically, a senior manager who is a key stakeholder in a critical evaluation can thus block career advancement for the officer responsible for that evaluation. The Panel notes this procedure creates a risk of abuse, but has not encountered verifiable evidence that this is part of practice.

To avoid this problem other organisations have evaluation as a separate career path or a ‘terminal position’. This is at variance with current practice whereby recruitment of OEDE staff strives to maintain a mix of experienced WFP generalists to help the department understand the unique nature and culture of the organisation and externally recruited evaluation professionals. The decision to increase the number of professional evaluators in OEDE is widely appreciated both as a way to increase performance quality and to reduce the number of staff subject to internal career pressures.

The Panel feels that on balance the mix between professional and career staff in OEDE is appropriate as long as the organisation ensures that career staff continue to be provided with ample opportunity for training and with adequate protection viz. undue pressure from Operations. In addition, only staff that express a strong desire to work in OEDE and who have relevant skill sets should be selected so as to avoid the notion that ‘anyone will do’.

Many WFP staff perceived the major problem of independence to lie in the lack of impartiality on the part of some external evaluators, who are seen to have preconceived ideas and concepts. This is sometimes seen as linked to weak understanding of the subject being evaluated or WFP’s policies and strategies more generally (e.g. preconceptions about the role of food aid).

Various concerns have been expressed that some consultants may be too close to the organisation, thus compromising their independence. Some are involved in evaluating and then subsequently assisting with subsequent formulation of a PRRO or other planning, or are otherwise contracted to carry out assignments related to the same programme. They do not sign a ‘declaration of conflicts of interests’. This is seen by OEDE as problematic, but not by others at Country Office levels and appears to be common practice regarding consultants engaged in decentralised and self-evaluation. Despite the advantages of using consultants with previously demonstrated competence and in-depth knowledge of a programme, the practice should be avoided within the same programme.

In the field, evaluation consultants and implementing partners report that they are able to discuss their experience without interference from senior management. There were indications, however, that the selection of field sites to be visited by evaluation teams may be left to WFP staff and implementing partners. Implementing partners privately acknowledged that they tend to show evaluators their best sites and activities in order not to risk compromising future contracts with WFP.

Concerning the participation of OEDE staff as part of the evaluation team (as well as the participation of Regional Bureaux staff on decentralised evaluations), the Panel feels that it is acceptable for staff to participate as members of evaluation teams as long as they do not per-

form the role of team leader (to preserve independence and objectivity). Their functions within the team (as resource persons), and reporting relationship with the team leader are clearly described in detail in the ToR.

The Panel concludes that the independence of OEDE-managed evaluations is largely safeguarded by OEDE evaluation managers, but that independence at times is compromised by “independent” evaluators who may have a vested interest in future contracts, noting that this is a problem that is not specific to WFP. Regarding decentralised evaluations, the Panel detects weaknesses regarding independence but feels that on balance, the priority focus of the Regional Bureaux and Country Offices should be to ensure a professional and objective exercise that can usefully contribute to better design and achievements of results.

3.4 Links between evaluation planning and budget

The Panel considers the link between evaluation planning and budgeting to be quite well developed and concludes that the Director of OEDE has sufficient independence in the selection of themes and operations to be evaluated. OEDE presents a biennial evaluation work plan that is budgeted concurrently. Nonetheless, the current budgetary cuts for WFP as a whole carry some risks affecting the independence of the evaluation function due to greater reliance on support from management and field offices for allocations. Initial budget discussions seemed to indicate a cut for OEDE – this was also based in part on the results of a senior management retreat which ranked OEDE fairly low in terms of corporate priorities. In the end, OEDE, unlike many other headquarters and Regional Bureaux escaped the general budget cuts virtually intact (including a slight increase in overall budget). However, an increase in staff costs and the increase in value of the Euro relative to the US dollar (USD) have resulted in less funds being available for non-staff costs of evaluations.

The allocation from PSA for USD 700,000 per year for evaluation activities in addition to USD 2 million per year in staff costs appears meagre, and constitutes in real terms a significant cut in comparison to the prior budget biennium. This is illustrated best when considering that current estimates budget about USD 190,000 per thematic evaluation and USD 120,000 per operations evaluation. This would effectively allow OEDE to implement about two thematic and three project evaluations, far shy of the planned 12 evaluations for 2008. However, this reduction is apparently off-set by another USD 3 million which is now firmly programmed for evaluations budgeted under DSC and non-PSA. If effectively implemented, this would constitute a sizeable overall allocation and commitment towards evaluation.

The Panel was informed that the Executive Director, above and beyond deciding to keep a strong evaluation function, has also made a

strong commitment to assure (a) that these funds are secured for evaluation use and (b) that these can be programmed for OEDE-managed evaluations. There is awareness on all sides that this should not result in a repeat of the budget situation in 2001/2002 when the entire non-thematic OEDE budget for evaluations was partially dependent on the interest and good will of Country Directors. At that time this dependency resulted in increased strained relations with Operations and the inability of OEDE to access some funds programmed for evaluations as they were no longer available by the time OEDE requested the funds. There were also reports of pressure by Country Directors on evaluation team members as the Country Offices had ultimate control over the funding. It is therefore of absolute importance for independence that these budgetary resources are carefully reserved for use by OEDE within the priorities of its work plan.

The Executive Board members who met with the Panel expressed a preference to become more involved in approving the strategic plan and budget of OEDE. While they currently have the opportunity for discussion when the Management Plan is presented, it is clear that they would like to become more involved. As earlier indicated, an option for doing so could be the creation of a Board sub-committee on evaluations (as it exists with IFAD) which could enhance the independence and utility of evaluations.

The Panel concludes that while the overall budget available to OEDE appears adequate to support the currently planned evaluation portfolio, the level of resources available could form a constraint for evaluation coverage and the scope of individual evaluations. The Panel recognizes that rough budget estimates are needed for programming purposes. However, the somewhat mechanistic estimation of evaluation budgets risks deepening the “cookie-cutter” approach used in the past where the cost and duration of an evaluation was programmed without consideration of evaluability, existing M&E data, size and scope of the intervention, and other special features. OEDE’s ongoing quality improvement processes should address and mitigate the risks inherent in this budgeting process.

3.5 Relationship between evaluation and audit

For a brief period (2001/2), OEDE was located in the Oversight Division before being moved to the Results-based Management Division (OEDR). Unlike the current weak follow-up on evaluation recommendations, follow-up of audit findings and recommendations is firmly secured in the organisation.

Some interviewees have suggested moving OEDE back into Oversight and Audit but few support this suggestion since many want OEDE to have a role as a ‘management tool’ rather than a perceived ‘policing’ or accountability role; it is feared that a shift would signal the latter.

Others suggest that OEDE should at least adopt audit-type procedures since it would then be in a stronger position to pressure for better quality management response. The Panel feels that a move to Oversight would not necessarily improve independence and that this could detract from utility of evaluation for management.

Both internal audit and evaluation have a combined function of accountability and learning. Audit judges that it balances the two functions roughly equally. Some field staff have complained that auditors are moving too much into programmatic issues without understanding them, while auditors have felt that some evaluations do not duly question existing data sets such as COMPAS. Evaluation focuses on broader policy issues and is therefore by nature more concerned with learning. At present, it is not always clear why certain activities are undertaken by OEDE and others by Audit (such as Audit's review of results-based reporting), although OEDE's undertaking of the *Evaluation of the Business Process Review* was apparently in response to a request from the Executive Board.

3.6 Ensuring access to needed information

No *structural* obstacles have been noted to access to available information. Deficiencies appear to stem from a lack of reliable information due to ineffective monitoring systems. There are also reports during highly sensitive evaluations of WFP managers obstructing evaluator's access to data.

3.7 Conclusions on independence

Overall conclusion

The Panel concludes that the Independence of the WFP evaluation function is adequate in comparison to similar organisations, but that the reported lack of respect for the evaluation function could erode this. Opportunities exist to exert pressure on OEDE staff due to human resource management structures and greater safeguards are therefore needed in this respect.

Conclusions regarding UNEG norms

Norm	Met?	Comment
Responsibility for evaluation	P	The Executive Board and Executive Director fulfil some of their responsibilities, in particular regarding the appointment of a professional head of evaluation, and, at present, safeguarding resources for evaluation. But the norms of clear accountability for the implementation of recommendations are not met and that of fostering an enabling environment for evaluation only partially.
Independence of evaluation	P	The evaluation function rates well in terms of overall independence. However, OEDE staff careers may be affected by their evaluation role, which is a considerable concern. There are insufficient safeguards to prevent partiality amongst external evaluators.
Reporting line	Y	OEDE is outside of line management while at the same time sufficiently integrated into WFP leadership structures to facilitate impact.
Conflict of interest	N	OEDE has not taken sufficient steps to ensure that evaluators avoid conflict of interests. The role of Regional Bureaux in both oversight and advisory support to Country Offices can have problematic implications for their role in decentralised evaluation.
Authorship	Y	Evaluation reports are issued in the names of the authors; however, the role of the OEDE evaluation manager either as author, quality manager or in guiding the team is not always clear.
Freedom of access/ conduct/ expression	Y	OEDE has full structural access to information and is free to express its findings.
Interested parties	P	Evaluations take the views of all relevant stakeholders into account, but the evaluation process does not provide for sufficient dialogue with stakeholders outside of WFP.

Y= yes P= partial N = no

4.

Credibility

Credibility is assessed in terms of both the evaluation products (reports) and the processes through which evaluations are managed. Furthermore, whereas the Panel has assessed WFP evaluations (both OEDE-led and decentralised) with regard to formal quality standards, it has been equally important to consider the perceptions of credibility within WFP and at different levels of the organisation. In some cases these perceptions are quite different from the Panel's assessment of quality. Some decentralised and self-evaluations are seen as highly credible within the countries where they have been conducted since they relate to the core concerns of the relevant stakeholders. Those evaluations have nonetheless been judged as unsatisfactory by the Panel and indeed by most stakeholders in Rome due to inherent faults in methods and a failure to adhere to basic quality standards. Perceptions of credibility are central to maintaining the respect that must underpin both independence and utility of evaluation within WFP. One of OEDE's greatest immediate concerns is undoubtedly that of restoring its credibility, which has been tarnished in the eyes of the Operations Division in particular.

4.1 Evaluation policy

OEDE's policy formation has included most of the elements of an appropriate evaluation policy. There are two challenges for the reader of the different evaluation policy documents. First, there has been a layering of policy statements, each containing important and for the most part highly relevant elements. It is difficult, however, to maintain an overview of what the current policy consists of. The official policy from 2003 has a well-considered perspective on independence, but the observer is left to infer (rightly or wrongly) that the aspects of credibility and utility that are rather weak in the document are addressed in earlier policies. It is impossible to determine which of these earlier positions is

still relevant, especially since the monitoring function has been detached from the OEDE, with possibly significant consequences for issues related to independence and how WFP addresses RBM in evaluation.

The second difficulty in interpreting the current documents is that most include a mix of policy statements, recommendations and general discussion of options. This leaves the reader uncertain as to which conclusions are confirmed and approved and which are not. WFP's self-assessment vis-à-vis UNEG's Quality Stamp provides additional examples of evaluation policy omissions.

The documents reviewed show an ongoing development of analytical and operational thinking with regard to how OEDE can better promote learning. Less attention has been given to accountability, apart from where it is directly related to internal WFP structures and reporting arrangements. Even there, apart from the accountability that can be expected to naturally flow from independence, there is little explicit attention to how evaluation can best ensure accountability. OEDE comments that 'For OEDE's understanding, accountability has to do with reporting on results – achievements and failures – that is in the sense of "giving an account of"'⁴. Accountability however includes action in response to evaluation findings, without which an account may be made but accountability remain absent. The development of mechanisms to ensure management action is limited to the collection of a written response for the Executive Board; follow-up and integration with management processes is largely outside OEDE's policy provision.

A major aspect of the policy perspective on credibility is the relationship that is envisaged with partners. Little is mentioned regarding learning and accountability among partners. Participation in ALNAP and this Peer Review process are the main references made to how broader credibility can be ensured. OEDE goals are almost entirely phrased in corporate terms, which by nature limit attention to accountability to external stakeholders, including the intended beneficiaries themselves.

In the relatively few instances where host governments and 'cooperating partners' are mentioned it is not quite clear what kind of relationship is intended and what role evaluation is expected to play in these relationships. Should evaluations treat them as objects to be assessed and ask whether these actors are accountable to WFP (i.e. following WFP policies and directives), or should evaluations play the opposite role by becoming part of a learning dialogue with these actors and enhance WFP's own accountability to them? Indeed, this may be a reflection of a lack of clarity on this issue within WFP as a whole.

In terms of use of the Evaluation Policy, most of those interviewed, including both WFP staff and evaluation teams, expressed either vague awareness of the policy or complete ignorance. Interviews suggested

⁴ Source: OEDE comments on the initial findings report August 2007

that attention to evaluation policy documents appears to have faded within WFP in the face of an overload of guidelines, positions and policies on different topics – not the least due to the prevailing confusion regarding the status of different plans, policies and guidelines for RBM, monitoring and related activities.

The Panel acknowledges that the various policies contain the major elements needed for a solid evaluation policy but judges that the layering of policies and limited attention to external stakeholders and accountability in its full sense, has affected its utility and effectiveness.

4.2 Monitoring and Evaluation Guidelines

The M&E Guidelines were prepared as an on-line tool for field offices when the two functions of monitoring and evaluation were co-located in OEDE. Clear directives were never issued regarding the status of the Guidelines after the separation of the two functions nor were they updated to reflect the new RBM system. They are considered to be a useful but somewhat cumbersome tool by many field staff. Awareness of their existence is considerably greater than for the evaluation policy itself. This is of notable importance given that a major factor limiting the quality of evaluation is the at times the low quality of monitoring and the confusion regarding the place of monitoring within the organisation.

There is, in effect, widespread uncertainty about (a) what the guidelines are for different forms of evaluation, (b) what they are for monitoring and (c) of the current status of documents that were developed under past structures. The M&E Guidelines are problematic – mostly as they are optional and there is no corporate policy that ensures compliance and a standard approach. Nonetheless the Guidelines at present fill an important gap within the context of a highly decentralised organisation with a weakly coordinated proliferation of mechanisms for monitoring and learning.

One ‘spin-off’ from the M&E Guidelines is the After Action Review (AAR) which was originally one of the options for self-evaluation in the M&E Guidelines. While initially co-piloting this approach with OEDE, Policy has further developed the AAR primarily by gradually simplifying the approach and reducing it in scope and ambition. A summary of the differences between AAR and evaluation has been prepared, but confusion still seems to exist. This needs to be addressed, but perhaps as part of a wider review of these and related mechanisms to avoid what could be perceived of as a piecemeal approach. The danger in the current situation is that if more guidelines are introduced without also formally discontinuing older guidelines this may be interpreted by the field as merely signalling that more work is being placed on the shoulders of their shrinking work force, thereby damaging the credibility of evaluation more generally.

The Panel feels that the lack of corporate guidance and policy on results and follow-up in general, combined with the proliferation of monitoring systems, self-evaluation approaches and other initiatives comes at significant cost to the organisation. The 2004 review of business processes and subsequent mapping of initiatives is an appropriate basis for forthcoming plans to clarify and streamline existing tools. Other corporate initiatives, including the WFP Information Network and Global System Phase Two (WINGS II) and OEDE's plans to clarify decentralised evaluation could go a long way towards streamlining data collection and evaluation tools.

4.3 Credibility of data used in evaluations

It is widely acknowledged within WFP that monitoring is fragmented, weak and often neglected. This raises significant concerns about the credibility of data used in evaluations given that most evaluations rely on existing monitoring data. A number of interviewees also commented that evaluation teams rely too heavily on anecdotal evidence. There are reports of rigorous monitoring systems being established and maintained by some Country Offices, sometimes in collaboration with local research institutes, but the extent of these efforts varies depending on the priorities and levels of commitment of the Country Office. In some cases, considerable monitoring data has been collected, but has not been analysed and used. Ultimately, the choice of data and judgements of its reliability are left at the discretion of the evaluation teams, and the advice of the evaluation managers.

Vulnerability Analysis Mapping (VAM) and the Commodity Movement Processing and Analysis System (COMPAS) are generally taken by evaluation team leaders as providing 'facts' or 'hard data' and not something that should be evaluated as well. Not questioning VAM /COMPAS data before using it to draw key conclusions hampers the credibility of the evaluation and the evaluation report. This 'practice' is of concern to internal audit which has, by definition, a critical view of the quality of data. Some evaluations accepting poor data at face value have indeed contradicted audit findings. This may be a result of the dearth of monitoring which encourages evaluation teams to grasp at any data, even if it is unreliable, since there is no other quantitative data available. Although there are exceptions, the Panel is concerned that this lack of reliable data is not reflected in the methods devised by evaluation teams for their short in-country missions. This has major implications for the credibility of their reports.

4.4 Competency and capacity of leadership, staff and consultants

There is a widely held view, shared by the Panel, that the overall competency and professionalism of OEDE staff is of good quality. All staff have availed themselves of professional evaluation training provided by ALNAP, IPDET or the Evaluators Institute and are continuously developing their skills set. The recent addition of professional evaluators (Director and one staff) seems to have greatly contributed to a perceived increase in credibility by Board and WFP staff alike.

Both the OEDE Director's and professional evaluator's posts were advertised externally in newspapers and, for the Director, via the UNEG Forum. P-4 Candidates were interviewed by the Director/Chief Evaluation Officer, and a Human Resources representative. For the position of the OEDE Director, candidates were interviewed by the Director of OEDR, the Chief and Senior Evaluation Officers, and selected Executive Staff. The P-4 candidate had to pass a written test.

There are two significant threats regarding OEDE's ability to maintain a high level of capacity in the short- to mid-term. The first threat is the extent of staff turnover which will occur within the department in the coming year, which coincides with severe staff reductions elsewhere in the organisation. The second related threat is the heavy workload planned by OEDE for the coming two years. Despite the training programmes available, it may not be possible to find time to train incoming staff without previous evaluation experience when there is a pressure to manage a large portfolio of evaluations.

Regarding evaluation teams, OEDE staff note major problems in recruiting quality consultant team leaders. This is partially (but not entirely) related to the relatively low level of UN rates for consultancy services. Currently the consultant register is being updated. Consultants' performance is assessed by the OEDE evaluation manager (an internal confidential grade 1–3 is assigned and a 1–2 page report is written). A grade 3 would typically mean s/he is not re-employed.

Several interviewees perceive an over-reliance by OEDE on a relatively limited pool of evaluators, which detracts from quality assurance, independence, a 'fresh' perspective and credibility in general in the eyes of field staff. The primary problem related to 'institutionalised' evaluators is their weak capacity to introduce new perspectives but also the impact on independence and perceived objectivity. However, figures from OEDE's database show adequate diversity, at least amongst team leaders,⁵ again indicating the need to manage perceptions in order to enhance credibility.

⁵ For example, over the past seven years only six team leaders have been used more than twice. (OEDE has used 46 different team leaders for 67 OEDE-managed evaluations, with 33 being used once, seven twice, four three times and two four times).

In a number of cases WFP staff complained that evaluation consultants arrived with pre-conceived notions and biases; in others, the evaluators lacked the necessary background in food aid which resulted in irrelevant recommendations. There is a general consensus that the lack of high quality, experienced and professional evaluation consultants is damaging the credibility of the evaluations and of OEDE.

In the view of the Panel, a contributing problem is the lack of a credible and transparent selection process of consultants. Further, an ongoing debate has been that of how to choose/mix team leaders and team members from different categories, i.e.

- OEDE staff
- consultants
- subject matter specialists
- evaluation specialists
- generalists with geographic competence
- generalists with WFP competence

The Panel supports the current practice of creating evaluation teams with a range of backgrounds that includes both OEDE staff and external consultants. However, it also recognises that the different roles and responsibilities of OEDE and external team members need to be better clarified. In that respect, the role of the leader of evaluation teams warrants special attention; he or she should be a very experienced evaluator capable of identifying bias, integrating the expertise of the individual team members and communicating the evaluation results in an appropriate manner.

It appears that the strengths and weaknesses in OEDE's competencies and its core consultants mirror, in many respects, those of WFP as a whole. The assessment of the Panel is positive regarding the evaluation of core WFP activities, but less positive with respect to crosscutting issues that the organisation has had difficulty addressing in recent years. Of these, increasing attention is now paid to gender, but human rights issues and protection are still addressed quite weakly. Notwithstanding some recent use of dedicated protection consultants and a greater focus on gender, the meta-evaluation for this Review notes that crosscutting issues such as protection and a rights-based framework remain weak.

4.5 Quality assurance

Quality assurance is the top priority of OEDE at present. There has been little systematic quality control until now and it was left to the evaluation manager to judge the overall quality and acceptability of an evaluation report. Since 2000, OEDE has, however, subjected itself to ALNAP's annual meta-evaluations, which have been fairly positive about the quality of WFP evaluations, in particular in comparison to

other agencies. Improving quality is seen by the new Director as a pre-condition for addressing other concerns and to counter unqualified criticism. Work is underway to clarify procedures, develop templates, provide staff orientation, plan training, etc. The draft guidelines presented to the Panel for review address many of the issues raised in this report, especially regarding clarity of purpose, stakeholder interactions before and during evaluations, and other core aspects of the evaluation process. Special attention is to be paid to the inception phase, which is well warranted for both ensuring quality and developing stakeholder buy-in.

4.6 Stakeholder consultation and engagement

While there seem to be significant gaps in terms of engaging external stakeholders during the planning and dissemination phases, evaluation reports and interviews indicate that stakeholder consultations have been reasonably well managed by evaluators within the constraints of time and funding. Time constraints in particular are at times so serious that questions can be raised about whether the means available to evaluation teams is sufficient for achieving stated aims. ToR are widely seen to be overambitious. Evaluations seem to be designed with for example, a standard allocation of time and approach to evaluation teams without considering the evaluation purpose, size of programme or country. As a result insufficient time may be spent at project sites and even less spent speaking with beneficiaries.

On the whole the quality of partnerships with respect to stakeholder engagement in evaluation practice reflects in part the ‘subcontracting’ nature of some of these relationships. Partners are not systematically consulted in ToR development, discussion of results and follow-up. These issues are reviewed in more detail below in the chapter on Utility.

4.7 Support to decentralised evaluations

As described in section 2.1.2, OEDE supports decentralised evaluation primarily through M&E focal points in the Regional Bureaux and is largely driven by demand, individual relationships and the degree of budgetary and human resources support to M&E in the Regional Bureaux. This is quite variable. OEDE staff provide only advice and no oversight or quality control. The quality assurance and other aspects of the OEDE reform process have not yet reached the stage of defining a quality strategy for improving decentralised evaluation, and this may inevitably await clarity regarding restructuring and budgetary decisions which will frame the ambition level. OEDE plans to develop its support strategy through consultation with the field.

Regional Bureaux have responsibility for overall oversight of the Country Offices and their engagement in decentralised evaluation is part of this task. However, few staff at this level have received appropriate training and the *de facto* responsibilities of the M&E focal point vary depending on the skills and background of the individual. A range of Regional Bureau staff are involved in evaluation and the M&E focal points usually are not in a position to train or fully support their colleagues.

In general there is a well-founded dissatisfaction within WFP regarding the focal point approach and little faith in the chances of fixing the M&E focal point approach unless it is institutionalised through a fulltime dedicated staff member at the Regional Bureau and a budget. There are even concerns that if such a measure was undertaken, staff and financial resources would be redirected to other activities due to the prerogatives of the Regional Bureau and Country Office Directors. Some of those interviewed also note that this is but one example of a range of tasks that are designed at headquarters and assigned to the field without looking at resource implications, incentives or direction regarding how to prioritise among these tasks.

The limited evaluation training that has been provided to Regional Bureau staff has not been seen as appropriate. Some Regional Bureau staff have requested evaluation training and noted that it is unlikely that the M&E focal point will be able to transfer this knowledge to colleagues. There are some concerns, however, that with the increasing work pressures on the Regional Bureaux due to budget cuts there may not be time for staff to attend training even if it was made available.

As there is no formal procedure for recruitment of an evaluation team, ToR development and implementation for decentralised evaluations, it is difficult to assess what 'normal' decentralised practice may be. Of those assessed, the approach is typically either recruitment through recommendation (nationally or from OEDE), a restricted sharing of the ToR with key stakeholders, and a shorter draft-comment-finalisation process (i.e. comments on the draft may not be sought from headquarters units commonly involved in the draft process of an OEDE report). Decentralised evaluation reports do not go to the Board nor are they shared on WFP's website. OEDE reviews all decentralised evaluation reports as part of its Annual Report analysis; it has found the quality weak and the credibility questionable. Including results from poor quality decentralised evaluations undermines the accuracy of the synthesis in the Annual Report, which in turn raises questions about the empirical basis for overall strategic decision-making in WFP if this report is seen to be a significant contribution to strategic planning.

There are reported incidences where decentralised evaluations are unduly influenced by Country Directors who are concerned that findings may disrupt existing plans. There is no process by which OEDE

can address this. The Regional Directors may have reason not to expend limited political capital protecting the independence or credibility of these evaluations.

Many interviewees attribute deficiencies in decentralised evaluation to the lack of competence in Regional Bureau to manage evaluations. Others see it as an issue related to the quality of the evaluation teams, which is in turn related to budgetary constraints. Since OEDE has larger financial resources, they are viewed as more likely to recruit better evaluation teams. Decentralised evaluations also make greater use of staff as evaluation team members or team leaders, which may compromise objectivity and independence.

To summarise, while many field staff felt that decentralised evaluations serve a useful learning purpose, their quality is variable and is often poor. They lack quality control and standards. At WFP headquarters, their credibility is low and not seen as particularly useful. Operations reported that from their perspective, clear proposals for improving or alternatively abandoning decentralised evaluation are urgently needed.

A number of reasons were cited for doubting whether decentralised evaluations could be improved, including:

- The field sees the task as having been “dumped” on them without sufficient human or financial resources to do a good job⁶,
- When evaluative tasks are reliant on limited project funds, this is one of the first items to be cut,
- Good quality evaluations are especially important now when programme design must be linked to higher level objectives such as PRSPs and UNDAFs, something which is seen to be beyond the capacity of Country Offices to critically assess,
- M&E focal point staff are frequently rotated and are usually not in a position to train or effectively support their Regional Bureau colleagues who are leading the majority of decentralised evaluations.

There are a number of decentralised evaluations planned for the coming two years but, based on experience in prior years, there is uncertainty about whether they will actually be budgeted for and implemented. If performance is to improve, senior management commitment will be essential and Country Office Directors will have to be held accountable if they do not undertake planned evaluations. If OEDE has to compete for those funds and is dependent on agreement by the Country Office, this may well result in debate (perhaps enlightening) over the relative value of OEDE versus Country Office- or Regional Bureau-led evaluations. This may be seen differently in each region and is closely related

⁶ In fact decentralised evaluations are not a shift in work load from HQ to the field, as OEDE's work programme remains more or less constant. It represents an intended overall increase in the number of evaluations conducted.

to the perceived utility and credibility of OEDE-managed evaluations. If senior management keeps up the pressure for increasing the number of evaluations some Country Offices and Regional Bureaux may ask for OEDE assistance simply since they presently do not have the human and financial resources to undertake evaluations themselves.

OEDE's current Management Plan includes proposals that deal with some of these challenges; others require broader and more concerted commitment. At the same time, there are drastic staff and budgetary cuts underway within the Regional Bureau, so the ultimate outcome of any capacity building effort at this stage is difficult to predict.

The Panel feels that decentralised evaluations must be retained to ensure broader evaluative coverage and utility and to allow OEDE to focus on more strategic evaluations as planned. Notwithstanding concerns about quality, the Panel supports suggestions that have been made to post them on the internal web site. However, more needs to be done to address quality and credibility concerns, and it will be incumbent on OEDE to deliver on its plans to develop appropriate tools, training programs and capacity development to ensure better quality control and ultimately good decentralised evaluations.

4.8 Role of self-evaluation and alternatives such as After Action Review

There are some suggestions, primarily put forth by the Policy Department, to abandon efforts to improve decentralised evaluation and revert to AARs (as a form of self-evaluation) instead. The lack of depth, independence and objectivity in AARs has made this a controversial proposal, which is generally not supported by OEDE and is questioned by many throughout WFP.

In essence, the difference between decentralised and self-evaluation is that the former seeks to retain a fair modicum of independence whereas the latter generally accepts that independence will be minimal and that it is more important to strive toward an acceptable degree of objectivity. Interviewees frequently did not clearly differentiate between decentralised and self-evaluation. This may be partly due to the active role of the Regional Bureau in supporting many self-evaluations and the active/direct involvement of Country Offices in Regional Bureau-led decentralised evaluation. There is a de facto continuum of approaches between the two, even if this has not been the intention. The Panel welcomes OEDE's plans to clarify and define concepts and terminology as a needed first step if additional resources are to be productively invested in self-evaluation.

In interviews some complaints were expressed about self-evaluations being cumbersome. AAR is beginning to replace self-evaluations in situations where the latter is seen as too difficult or costly. There is little awareness among most staff that the AAR actually originated as a

self-evaluation tool. When self-evaluation was created in 2002 it was to replace the final project report through more active dialogue on outcomes. The main rationale was to get more evaluative feedback and performance information beyond the limited information that was usually being fed into WFP's reporting system. Apparently it is sometimes felt that this is too difficult. The suggestions for replacing self-evaluation with AAR can be perceived of as an acceptance of weakness, rather than a way of addressing the limitations uncovered in the self-evaluation process. Over the past few years AAR has shifted from being designed as an ambitious and broad stakeholder consultation (e.g., pilot efforts such as the Georgia AAR) to a largely internal reflection activity that is not necessarily intended to reflect critically on overarching issues such as relevance. If beneficiary and other field level stakeholders' inputs in the AAR are not 'relevant', this indicates that questions considering the overall appropriateness of the programme to the beneficiaries are not considered relevant. This is of notable concern since AARs are intended to feed into the strategic planning process. If wider strategies are not questioned, and these activities are only undertaken to consider minor programme modifications, this may not even constitute strategic planning per se.

Sometimes it is reported that the choice is not between self-evaluation and AAR, but rather between AAR and no assessment of past performance at all since there are no resources available for evaluation. Others note that even the modest AAR approaches currently being promoted have cost implications. A range of views exists on these choices, partially since the nature of the AAR concept is still somewhat ambiguous.

In the view of the Panel, AAR is a useful self-assessment tool but cannot replace an external and objective evaluation. More clarity is needed in the terminology and in the policy. One way of clarifying the confusion may be to better anchor AAR (and other variations of self-evaluation) as part of the annual strategic planning cycles undertaken at the project level, rather than labelling this an "evaluative" activity.

4.9 Basic criteria for quality of evaluation

The Panel assessed the quality of evaluation reports based on a modified version of the ALNAP Quality Proforma (which is based on the UNEG norms and standards) and on 17 of ALNAP's own meta-evaluations. The main assessment headings were:

- the ToR
- methodology of the evaluation
- assessment of the intervention
- cross-cutting issues

- assessment of the report
- overall comments

To the Proforma has been added the assessments of Peer Panel members as well as interviews conducted with team leaders, evaluation managers and field staff.

The presentation of methodology in the majority of evaluations reviewed is weak. In some cases, a description is provided in the annexes; in others its absence makes it impossible to accurately judge other aspects of their quality.

While there are cases where WFP staff have involved partners systematically throughout the evaluation process, from consulting on draft ToR to organising workshops for dissemination and reflection, this seems to be the exception rather than the rule. Most WFP evaluations involve partners in an *ad hoc* manner and NGO cooperating partners generally behave similarly towards WFP, even though they are evaluating projects that depend heavily on WFP resources. This situation detracts from learning and ownership of results. It also undermines the credibility of WFP evaluations since ToR do not adequately reflect partner perspectives and results are often perceived as WFP-centric and lacking wider contextual relevance. The Southern Africa PRRO was one example observed during the current Peer Review where the evaluation was not seen to be answering critical questions regarding WFP's current value-added in southern Africa in the absence of a large emergency. Whilst this perspective was certainly demanded in the evaluation's ToR, it was not backed with appropriate resources. A similar situation was noted in Indonesia.

Overall, both OEDE and other WFP staff at headquarters acknowledge a varying degree of credibility in OEDE products and a rather low level in decentralised evaluation.

A major challenge for OEDE is how to deal with evaluations that are not of an acceptable standard but must be 'salvaged' due to the investment that has been made and the expectations that have been raised. Although rare, OEDE staff acknowledge discomfort when faced with very poor quality evaluations which have to be rewritten.

The impression gained from discussions with the Executive Board is that they consider evaluations credible, although they note the lack of hard data to support some reports and express some concerns about methods. It also appears that the professional credentials of the new Director have raised credibility. The Executive Board has a basic belief in evaluations and has recently used the evaluation management response to increase its pressure on management by requesting an improved and more coherent response.

4.10 Credibility of recommendations

Among a range of stakeholders interviewed the major concerns regarding quality consist of the following:

- unrealistic and un-prioritised recommendations,
- evaluators seen to be ignorant of the financial constraints of the operations and the financial implications of their recommendations,

Opinions differ as to whether assessment of the financial feasibility of recommendations is the concern of evaluation or whether it relates to management responsibilities. The role of evaluation is to highlight concerns and failures to achieve stated objectives but not to replace management's responsibilities to prioritise and define goals in a realistic manner. This difference of perspectives may come to the fore even more in the future given that the gap is growing between the very ambitious objectives presented in logframes and the financial realities of declining donor support to WFP.

Despite these concerns, it is not self evident that concerns about the quality of evaluations always limit impact. A lively debate may result in significant impact even if the quality of the evaluation is seen as being low. Also, a weak evaluation on an important topic may have impact despite its shortcomings. For example, Policy sees the Targeting Evaluation as being particularly influential, whereas OEDE (and some of the evaluation team itself) sees it as being of very poor quality. This raises of course concern about selective or biased use of evaluations and the possible distortions of results from operations.

4.11 Conclusions on credibility

Overall conclusion

The Panel concludes that the Credibility of products of the WFP evaluation function is uneven, and that the process of the function is somewhat more credible but also problematic.

Conclusions regarding UNEG norms

Norm	Met?	Comments
An official policy on evaluation	P	An official policy exists, but its nature as a layered series of documents detracts from clarity and applicability. Evaluation policy is not sufficiently used to guide practice. Evaluators and Regional Bureaux have not been sufficiently aware of OEDE expectations regarding quality standards due to a lack of specification within OEDE itself.
Evaluation is impartial	Y	OEDE is largely seen as impartial, but there are claims that some hired evaluators have preconceived ideas and there are cases of conflicts of interest. Although the views of all stakeholders are often sought, there is an uneven emphasis on stakeholders who are more accessible and articulate, with beneficiary views in particular under represented.
Evaluability is assessed	P	Performance is reasonable in this respect but ToR are overly standardised, over-ambitious and not always suited to the needs of the task. This is partially related to weaknesses in overall programme design wherein goals are not clear and logframes are often ignored. There is a dearth of monitoring data in most cases and existing data (e.g., VAM) is accepted without sufficient scrutiny.
Quality of evaluation	P	The quality of evaluations is mixed. This may improve with the establishment of new quality standards, but these are not in place at this time. Particular weaknesses have been noted in methodology and in some crosscutting issues such as human rights, where WFP as whole lacks capacity and guidance. A failure to take into account the cost implications of recommendations, together with factors related to the nature of priority setting in WFP has damaged the credibility of evaluations.
Quality of partnership	P	Partners are engaged primarily as either hosts or key informants, which is inappropriate for an organisation that should be encouraging two-way accountability and learning. The extent to which stakeholders are consulted in the planning, design, conduct and follow-up of evaluations (UNEG standard 3.11) is patchy and overly concentrated on the implementation stage.

Y= yes P= partial N = no

5.

Utility

The overall use of OEDE-managed evaluations is institutionalised in three main ways. One is that evaluations are made available to the Executive Board in connection with discussions on new operations. Likewise thematic evaluations, although not connected to individual programs, are presented and discussed in the Board and, when appropriate, linked to the presentation of a new or updated policy on the same subject. Another use is that evaluations are an important input to the Programme Review Committee prior to approval of new operations. A third institutionalised use is the management response matrix.

UNEG norms of intentionality stipulate that evaluations should be timed in such a way as to coincide with programme and budget decision-making cycles. The Panel found that programme evaluations were generally intended to coincide with Country Office programme decisions. Thematic evaluations, such as Emergency School Feeding, are timed where possible to influence or contribute to policy development. The presentation of OEDE's evaluations to the Executive Board suggests an intention to embed evaluation systematically in WFP's principal governance mechanism.

Many field staff tend to value evaluation for its contribution to programming decisions. All WFP interviewees felt that mid-term country programme and PRRO evaluations were useful to some degree as they influenced the design of the next phase. There seems to be a clear preference to see the main utility of evaluations as a stage in the formulation process rather than an accountability tool.

The Panel has not tried to chart the use of individual evaluations but the interviews show that particularly some recent thematic evaluations, e.g. on Emergency School Feeding and SENAIP, have obviously been widely read and used.

5.1 Purpose of evaluation

The current evaluation policy and the ambitious new Management Plan largely clarify the purpose of evaluation within WFP. However in order to more firmly explain the potential use of evaluation, a long term corporate evaluation strategy is needed to position evaluation and its dual function for accountability and learning firmly within WFP. The Management Plan for the next biennium is an effort to develop a more strategic approach to evaluations but does not explore how to ensure greater influence of the evaluation function on corporate learning and evaluation use. The Plan contains a number of assumptions that will have to be carefully tested. The dimensions of how the evaluation function will feed learning back into WFP and to the broader community (other than a passing reference to ALNAP and UNEG) is not made clear.

Past efforts to define the purpose of evaluation have involved promotion of both accountability and learning. On the one hand there have been efforts to instil an acceptance that WFP must be held to account for its work. On the other has been the need to ‘sell’ evaluation as a support function for programming and as an integral process of obtaining support for the next phase (especially for PRROs). ‘Learning’ is used quite broadly in policy documents, without a clear definition of what it means with respect to organisational or individual process and is often characterised by learning what to put in the next programme document.

Tension exists between the learning and accountability functions. The term accountability is perceived by many within WFP as an activity closely related to audit and as being addressed through the submission of all OEDE-led evaluations to the Board. This also reflects a limited and rather internalised interpretation of accountability. In the view of the Panel there is clearly an imbalance between the learning and accountability function with the former being the ‘desirable’ but clearly underachieved outcome and the latter being the current ‘perceived’ outcome. Neither functions to full potential and the future evaluation strategy should detail how both functions can be achieved simultaneously, be better balanced or demonstrate an explicit prioritisation of function.

5.2 Uses of evaluation for programme decision-making

The Programme Review Committee (PRC) process is seen as one of the main channels to use evaluation for impacting on programme decision-making. Comments received indicate that the PRC process is seen to be essential but flawed. Concerns about the influence of OEDE on the PRC include the following:

- the PRC does not always post evaluations as background documents for design of new programming,

- the PRC takes place late in the process of programme design and decisions have often already been made regarding programme plans before OEDE is asked to comment, leaving Country Offices with insufficient time to make significant changes, and
- there are few repercussions for ignoring PRC recommended changes due to weak follow-up and its advisory status⁷.

Suggestions have been floated that these issues should be addressed by the Internal Evaluation Committee (IEC). Other ideas have been put forth that the weakness of the PRC process be addressed within the overall restructuring as part of a new overall Design, Monitoring and Evaluation structure with more explicit attention to the role of log-frame analysis.

There are some concerns that WFP is not using evaluation to highlight the fundamental programme challenges facing the organisation. Evaluations are seen by some as part of public relations, concerned with ‘tweaking’ existing programming (asking whether WFP is ‘doing it right’) rather than asking whether WFP is ‘doing the right thing’. One reason for this is that general problems have been noted with attribution issues in evaluation, i.e., how to say whether food aid solved a problem and how to prove whether or not food was an appropriate solution. This is especially problematic with Food for Assets and Food for Work since outcomes/impacts are related to factors far beyond the remit of WFP.

Evaluation has not contributed significantly to situating WFP in the discourse on food aid versus cash-based responses, or on the role of humanitarian aid modalities in middle-income countries. WFP pays very little attention to non-WFP evaluations since these are not seen to be relevant (and staff are already overwhelmed by documents). These areas of accountability beyond one’s own project portfolio have been a major aspect of how the Humanitarian Accountability Partnership has grappled with the accountability of the sector. Some lessons may be drawn from that experience.

It should also be noted that in some cases evaluation is used to justify politically difficult shifts in terms of coverage or even of a more fundamental nature (e.g., cash versus food). Also, accountability to *current* partners may sometimes be seen as limiting the ability to take a broader perspective on the possible need for a radically *different* set of partners. Feedback indicates that evaluation has sometimes played a positive role in the narrower aspects of this (e.g., geographic shifts and targeting) but has usually not taken a sufficiently ‘out of the box’ perspective to question the basic choice of priorities in a country portfolio.

⁷ A study conducted by OEDE in 2005 showed that, over the period 2004–2005, only 36% of OEDE’s PRC comments were responded to by Operations.

Some express concerns that evaluation has generally insufficient impact on programming, primarily because programming is not evaluated based on logframes, policy commitments, etc. OEDE has signalled a strong commitment to evaluate based on logical framework analysis. This would seem to be a tool to promote utility among Country Office staff since it would tie evaluation closer to the programming framework. Problems arise in that Country Offices may have an ambivalent attitude toward logframes as a tool for programme design and monitoring which stems from the broader confusion regarding RBM in the organisation. They have apparently received mixed signals regarding the importance of logframes, particularly regarding expectations for how they will be applied in practice. The decision to apply them in evaluation, while fully warranted in many respects, may lead to criticisms of lack of direction in country programming when the failure to follow logframes may be due to a lack of leadership and strategic direction at higher levels.

An example of the response to this conundrum is that the reports of the evaluations of the Ethiopia and Kenya PRROs (programmes totalling close to USD 1 billion), which were conducted at the beginning of the year, are being extensively redone based on logframes and stated objectives. These are seen as good examples for raising expectations regarding evaluation quality since these operations are so well funded that budgetary constraints cannot be claimed as a reason for not focusing on overall aims. This is creating significant tensions since field offices have not expected to be evaluated based on logframes. It is too early to draw conclusions about how this seemingly obvious (but apparently controversial) approach will impact on utility.

5.3 Meeting the needs of different users of evaluations

Just as there is a range of uses of evaluations in WFP, there are also multiple users. Yet the Terms of Reference tend not to tap the full potential of other users of the evaluations. Draft standards for OEDE state that the Executive Board is the “primary user” of all evaluations, but it is clear that there are intentions that a far broader range of stakeholders are expected to benefit from their products. Interviews indicate that staff in Operations are tactically oriented and primarily want to learn about specific programmes and have information about performance to meet donor reporting demands. The Executive Board is more strategically oriented and wants to draw broad lessons about where WFP is or should be headed. The Executive Board also sees evaluation as a tool to promote broader learning in the organisation as a whole and sees its own role as acting as an advocate for increasing the utility of evaluation within WFP. Country Offices have a range of interests. It is in many respects these demands, rather than the nature of evaluations supplied, which frames utility.

Regarding external potential users, with some notable exceptions, WFP's broader accountability to partners can be characterised as falling within the following continuum:

- For some it is a non-issue, as partners are seen as either a threat to neutrality or an unavoidable burden.
- In some cases the relations with partners are strained, and discussion of accountability toward them as part of a more inclusive approach to the evaluation process could make this more difficult and perhaps weaken WFP's negotiating position.
- Some partners are seen to be actively trying to influence evaluation findings to promote the vested interests of their organisation, which is notably problematic when these partners' staff are members of the evaluation team (e.g., Indonesia PRRO).
- In others, evaluation is seen as one aspect of an ongoing process of building ownership for activities within counterpart structures, with accountability perhaps being an ideal aspect of this.

Another factor impinging on the limited accountability to stakeholders outside of WFP is that evaluation has focused more on internal WFP process than programme outcomes. Some outside stakeholders report that they are not very interested in WFP's own internal processes (apart from those that are related to sub-contracting and partnership), and thus have limited interest in many aspects of WFP evaluations. The nature of decentralisation in some countries and the operational nature of WFP work have meant that in many instances NGO partners in the capital pay less attention to WFP's work. The 'real' partners are in the field and the distance means that their engagement in a capital city debriefing presentation may not be a realistic goal. Moreover a failure to engage partners in the evaluation process does not necessarily indicate an overall failure to engage in terms of accountability and learning.

In general, approaches to wider consultation and transparency are somewhat more constructive with regard to governmental counterparts than to NGOs, and in turn more oriented toward NGOs than the private sector. WFP rarely uses evaluation as a way of encouraging broader accountabilities in respect to its position within the humanitarian (and sometimes development) community in a given country or even internationally apart from participation in ALNAP. NGOs do participate in annual consultations with WFP but evaluation is only mentioned in passing in these meetings, which have thus not been used as a forum for dialogue on evaluation findings and recommendations.

The private sector is a major stakeholder in many WFP programmes as a transporter, biscuit manufacturer, etc., but since they are openly seen as subcontractors, the idea of engaging them as a 'partner' in the evaluation process has not been raised. The private sector is also

increasingly acting as a donor to WFP (USD 8 million this year in Indonesia) and may be influencing WFP priorities accordingly, but this issue has not been explored as either an evaluation topic or as a new aspect of accountability mapping.

It would seem that OEDE could bring the needs of different users together through raising attention to the broader implications of findings from different evaluations. This could serve both those looking for wider trends and those concerned more with how to improve a specific programme. OEDE is seen to be making some initial progress toward this, but much remains to be done.

The Panel believes that better stakeholder mapping and a more appropriate communication strategy would ensure there is greater ownership and improved learning amongst internal and external stakeholders. Such a communication strategy would need to consider how different stakeholder groups communicate and learn and whether resources need to be reallocated accordingly.

5.4 The management and follow-up of evaluations

The range of types of evaluations and the consequent variety of constellations of key stakeholders for learning and accountability have meant that OEDE has in the past not been able to adopt a standard approach to managing evaluations aimed at both learning and accountability. Much has depended on the ingenuity of the evaluation manager, the openness of stakeholders to engage and the mutual trust among all parties.

Stakeholders report being confused regarding where to concentrate their engagement when similar reviews and studies are being conducted by different departments within WFP. With regard to thematic evaluations, for example, Policy has sometimes had related initiatives underway parallel to OEDE evaluations on similar topics and coordination has not always been effective.

The lion's share of an evaluation process is spent on the evaluation's middle-phase of implementation. Pre-evaluation inception missions, where conducted, have enhanced the preparation of the evaluation and are widely appreciated, improving focus and evaluability. However, relatively little effort is spent on canvassing external stakeholders' perspectives on design and utilisation or otherwise preparing stakeholders for evaluations. This in part leads to a lack of buy-in or ownership for the evaluation and its results. There are some notable examples of good practices, e.g. the SENAIP thematic evaluation.

Regarding follow-up for learning, efforts are not well structured. Some positive *ad hoc* initiatives have taken place (e.g. the follow-up workshop on the Emergency School Feeding evaluation), but these are the exception. On the whole, once the evaluation report has been submitted to the Executive Board, there is little additional effort to get evaluation results known and used throughout the organisation. Neither reports

nor executive summaries are translated into languages other than official UN languages, even though this would be essential for wider consultation and transparency in some countries. There is no active process of drawing the attention of stakeholders to the final report or engaging in further follow-up after the debriefing. In the view of the Panel this severely affects the utility of evaluations as spin-offs such as short publications, presentations around the organisations, direct feed into training, special follow-up workshops etc. are known to be effective tools to promote evaluation use. Good practice seems to be more related to the initiative of the Country Office than to OEDE procedures.

Since OEDE has no follow-up or dissemination strategy other than the presentation to the Executive Board, there are many missed opportunities. For example, if an evaluation team leader is invited to the Executive Board presentation, adding a (half-day) workshop as was done for the Emergency School Feeding session is not very costly and can serve to better disseminate the evaluation and trigger internal debate. Nonetheless, dissemination at field level does remain a cost issue that should not be underestimated.

The Panel finds that utilisation is likely to be more effective if there is a clearer view and strategy on dissemination following the evaluation process itself. A practical way to enhance dissemination would be to include this in the ToR and in the budget for any evaluation. It is the Panel's view that OEDE's planned increased use of stakeholder workshops in the field (during the evaluation and post evaluation) could be an important contribution to increased dialogue, evaluation quality and buy-in.

5.5 Joint and inter-agency real-time evaluation

Joint evaluations and Interagency Real Time Evaluation (IA-RTE) are being discussed but at the moment are not a prominent part of WFP evaluations planning. WFP has engaged with joint (FAO, UNHCR) and inter-agency RTE in the past but has not made this a priority nor has it been part of a deliberate strategy. Like other organisations participating in inter-agency evaluations, the follow-up is weak, which may reflect in part lack of appropriate internal buy-in to such exercises. According to agencies interviewed by the Panel, WFP, while involved, has not played a major role to date in the inter-agency RTE pilot. Partner agencies felt that a stronger engagement of OEDE in the 2007 Mozambique and Pakistan IA-RTE's would have been warranted, in particular given WFP's role during the response. In both evaluations there was engagement by WFP Country Offices, and to a certain extent by headquarters, but the role of OEDE was seen as not significant nor did WFP share any of the central costs. WFP's strongest engagement in IA-RTE has been the Pakistan cluster RTE in 2006 when OEDE funded the evaluation team leader.

Joint evaluations with FAO and/or IFAD could add a broader and more informed perspective on links between relief, rehabilitation and development programming and even on issues related to the complex nature of risk reduction (insurance, food security – agricultural commercialisation trade-offs, etc.), which are sometimes glossed over in discussions of the Hyogo Framework. Talks are now under way with FAO about a joint evaluation of food security information systems, which would become part of the strategic evaluation on emergency preparedness and response. This demonstrates the intention of OEDE to capitalize more on such partnerships in the future. Joint evaluations with UNICEF or WHO could similarly provide a greater perspective on nutrition, while joint evaluations with UNHCR could focus on the impact of food aid on the food security of refugees. The past good collaboration with UNHCR has come to a halt with the temporary demise of UNHCR's evaluation function.

OEDE has indicated to the Panel that they intend to focus less on agency-specific RTEs and to make full use of the inter-agency RTE mechanism. While there seems to be little demand for RTE (from both the Executive Board and the organisation), the Panel suggests that some flexibility may need to be retained to give more timely and relevant feedback to emergency programmes. It would be useful to provide space in the work plan for more immediate evaluations – and in particular those that are not necessarily linked to the formulation of a new programme. Regarding inter-agency RTE, the indication is that a stronger commitment and internal advocacy is needed for them to be seen as useful for the organisation. As a sign of the weak ownership of joint initiatives, WFP did not undertake a management response to the evaluation report of the Tsunami Evaluation Coalition (TEC) although OEDE was a member of the TEC.

The Panel discussed an idea which emanated from ideas surrounding the “One UN” concept of merging the evaluation units of the three Rome agencies but felt that given the present need to enhance the internal effectiveness and credibility of evaluation within the organisation, there seems little value in moving OEDE from its location within the organisation. Furthermore the Panel felt that OEDE's ability to potentially influence strategic thinking and learning within the organisation could be adversely affected.

5.6 Contribution of evaluation to managing for results

Although OEDE was formerly placed in the RBM division (OEDR), evaluation during that period was separate and did not appear to significantly influence the development of the corporate RBM system. Nor has the corporate RBM system significantly influenced the results-focus of evaluations. Evaluation findings and conclusions are used for the annual performance report of WFP but do not provide consistent or

user-friendly information on results. The Panel noted that many reports focus mainly on input, activity and output information but that few evaluations provide solid performance information at outcome level. As a consequence the use of evaluations in the annual report appears anecdotal.

When considering the use of evaluations for “managing for results” the Panel found that evaluations are indeed used for decision-making for future programming but seem to have little influence over corporate decision-making and results-based management. The reason for this is the lack of good monitoring data which is linked to the evaluations that are not seen as providing credible results-information beyond the goods and services delivered (outputs). It is therefore not surprising that there appear to be few consequences for operations whose evaluations have not been able to demonstrate any tangible results or benefits. Given the above, it would be unrealistic to use evaluations for “managing for results”. The Panel feels that data quality and the overall quality of evaluations and their ability to assess results has to be improved before evaluation can fully play the role it is intended to within a results-based management system.

The lack of a consistent approach to field-based monitoring also affects the quality of evaluations and RBM. This seems to be a perennial problem and has not been addressed during the years of OEDR. At the time of the Panel visit, there was no ‘corporate home’ for monitoring or RBM for that matter (with the exception of one lone RBM officer), nor could the Panel discern the presence of a corporate monitoring strategy that is applied by each Country Office. Although there have been significant efforts at headquarters to define output and outcome indicators, it is largely left up to each office to put these into practice and design its own M&E programme. Some of that information is fed into the annual reporting but the lack of consistency between systems raises questions regarding data credibility.

There are reports of very comprehensive country-level M&E systems (e.g. Sudan) but in general systems very much depend on the initiative and ability to resource these by Country Offices. Allowing such proliferation and fragmentation of M&E systems also risks duplication and is costly to the organisation. The role of evaluation within RBM has also not been firmly established. The M&E Guidelines that were created by OEDE in 2002 (when OEDE had a monitoring officer) are still in use but have not been updated to take into account the evolution of the RBM approach between 2003 and 2006.

Another potentially relevant tool was the Common Monitoring and Evaluation Approach (CMEA). The CMEA was intended as a key component of the new RBM strategy, devising a set of common corporate objectives and key performance indicators that could be tracked centrally. A by-product of the CMEA was to include a ‘paper trail’ for

evaluators and concise results information. Developed as a stand-alone programme, the CMEA was eventually intended to become an integral part of the corporate information management system, WINGS. However, WFP staff familiar with the CMEA expressed doubts that this system would ever become operational, in part due to the fact that the organisation has run out of funding for the system, in part due to difficulties making it compatible with new WINGS-2. If CMEA survives – at a recent WINGS-2 meeting it was confirmed that some funding was available to continue its work – it will be limited to produce output information. This begs the question of how WFP will ever capture outcome information.

A recent internal self-assessment on the implementation of RBM in WFP issued by the Office of Change Management in October 2007 acknowledges that while good progress towards establishing coherent results-based management and long-term objectives has been made, RBM has not been adopted by all offices and that there is a need to develop effective performance information and monitoring systems to better make use of their findings. Only about 30 percent of mainstreaming RBM activities had been completed at the time of this report and there remains the pressing need to define appropriate results chains, indicators and targets in most planning and project documents. This data confirms that RBM has been mainstreamed too early and too fast. The Panel feels that unless there is continued investment in rolling out all features of the planned RBM system, the good initial work done on RBM will most likely be wasted and result in a sunk cost for the organisation. Furthermore the Panel feels that before proceeding further with the development of the tool, an independent evaluation of RBM is called for to help assess objectively the future of RBM in WFP.

The Panel concludes that the impediments of the current RBM and monitoring system seriously hamper the ability of evaluations to provide and analyse results information. It appears to the Panel that the major investment in RBM has been a missed opportunity for the organisation to develop a more consistent and systemic approach to results monitoring, reporting and evaluation.

5.7 Contribution of evaluation to policy making and organisational learning

Efforts are under way to improve the learning dimension of evaluation but it is so far not fully utilised. There is limited ‘harvesting’ of evaluations for applying key evaluation insights in decision making. Evaluations do not reflect lessons from earlier reviews, nor are those lessons systematically collected at a meta level and analysed. Thematic evaluations do address this but they are not fed (nor are intervention-specific evaluations fed) into a corporate knowledge management system. The current community-of-practice developed by Policy (PASSiton) bypasses evalu-

ation knowledge as its lessons are only drawn from recommendations made in AARs and from direct inputs by field staff. WFP, in the 1990s, had developed an evaluation data based called the Evaluation Memory System (EMSYST) that was collecting lessons. This system has been discontinued and apparently was barely used by OEDE. Provided it is clearly reflecting corporate, rather than individual-project learning, OEDE could seek to better utilise the PASSiton platform in order to have a stronger place in organisational learning.

Thematic evaluations are OEDE's main tool for using evaluation to stimulate cross-learning across regions. However, beyond making evaluations available on the website there is little attention paid at headquarters to proactively using evaluation to promote learning, either between different WFP Country Offices or Regional Bureaux, or with partners. Within regions this is normally the responsibility of the Regional Bureau but again, there is no evidence of strategies for doing this.

Thematic evaluations are intended to contribute to this more strategic and policy relevant thinking and indeed play an important role in raising the level of discourse above programmatic 'tweaking'. With some notable exceptions, identifying the use of thematic evaluations for such broader analysis has proved problematic as most interviewees have difficulty citing positive examples of evaluations that have concretely changed their thinking on these issues. This challenge, however, is by no means unique to WFP.

By contrast, the Executive Board gives much higher priority to learning from thematic evaluations as this gives them insight into understanding the effectiveness of what the organisation as a whole is doing and enables them to pressure for policies that reflect what WFP should be doing in the future. Related to this, thematic evaluations also help them to determine whether or not support should go to WFP for certain types of activities. This was noted by one Executive Board member as a major role for evaluation from their perspective. Programme and country evaluations provide useful snapshots of portfolios in practice (notably the Niger evaluation), but do not focus as much on the core policy questions facing the Executive Board. Board members expressed the view that the proposed ratio for 2008/9 between thematic and country-specific evaluations is too biased towards operational evaluations. They would also prefer more synthesis of the findings of programme and country evaluations, i.e., the 'harvesting' of findings from the evaluation portfolio.

In relation to the issue of evaluation informing decisions about whether or not WFP is the best agency to respond to core needs, the position of the Executive Board members, who are in many cases also on the Executive Board of FAO and IFAD, has also sensitised them to the need for WFP to better place its analyses within a broader context

of food security and even poverty. The belief within WFP management that OEDE should function as a ‘management tool’ may detract from looking objectively at such core issues. This potential goal conflict in the work of OEDE does not appear to be directly confronted as part of current reconsideration of OEDE’s strategy.

The Executive Board also indicated they wished to be consulted more on the selection of themes. One Board member said that when evaluations are presented during the regional discussions, back-to-back with proposals for the next phase of the programmes that have been evaluated, the results is less attention being given to the evaluation.

Some concerns have been expressed within WFP that evaluation has failed to identify a role for itself in helping the organisation to develop policies that address more megatrends, e.g. the debate on cash, normative versus operational roles, whether or not to retain a role in development, etc. (there is also a concern that Policy is not filling the gap in this regard either). There is some discussion of creating a Policy Review Committee to which OEDE would feed information and act as an advisor. But it is unclear if this would make a significant impact if OEDE already has a set work plan and if this is ‘just another committee’ to report to. Respondents generally indicate that some progress is being made toward looking at more fundamental issues, but there are a range of levels of satisfaction with this progress, presumably related to whether the respondents themselves wish the organisation to undergo more ‘soul searching’.

A recent decision to emphasise strategic evaluations may indicate a desire to better position evaluation with regard to policy, but the trade-off could be less perceived relevance and usability from the field if the focus is on feeding into the distant strategic planning process (and also being seen to be part of the proliferation of uncoordinated learning and monitoring initiatives in headquarters). There is already a significant frustration at the inflation of pressures from headquarters at the same time as budgets are being drastically cut in the Regional Bureaux, and problems may arise if OEDE is seen as part of this process rather than acting as a reality check.

5.8 Contribution of evaluation to knowledge management

OEDE posts all its evaluation summaries and reports on its external website. Some field staff point out that they cannot easily find information that is of value to them on specific issues. It is not easy for external users to find the evaluations on the external website.

An individual who is proactive and knows how to navigate the intranet can obtain copies of Executive Board transcripts, full reports, etc., but the current structure is not supportive of this and it is doubtful that this occurs very often. Similarly, decentralised and self-evaluations are not easily accessed as they are not posted on the website. In general

the current structures of access do not contribute enough to supporting the ‘culture of evaluation’ within WFP nor with its partners.

5.9 Management response to evaluations

A management response mechanism is established within the organisation and is one of the means to ensure use of evaluations. However, there is room for more effective use of this both for operations related decision-making and for reflection and long-term learning.

The management response matrix for evaluation is seen as a useful tool by both WFP and the Executive Board. However, there is insufficient clarity regarding the ownership of the management response and the follow-up of the management response is *ad hoc*.

A WFP study on WFP’s follow-up to evaluation recommendations notes that staff’s assessment of utilisation was ‘intuitive’ and not always reliable⁸. Information provided from Country Offices and Regional Bureau to OEDE for its Annual Report regarding implementation of recommendations is admitted to be subjective and unverified. It is based on the recommendations that have been marked as ‘accepted’ or partially accepted, not on whether any action took place.

The management response is a fairly blunt tool that works better for the field of internal audit from where this tool was adapted. The interaction between evaluation recommendations and management follow-up as expressed through the matrix is overly mechanistic. Management response is not simply a matter of showing a high degree of compliance with recommendations but rather an opportunity to critically reflect on recommendations and suggested approaches. More often than not the management response has not constituted the dialogue between management and evaluation it was intended to be. The management response has apparently gained in prominence since the Executive Board has started to question some of the management responses.

On the whole it can be said that WFP’s mechanism for management response compliance is weak. The monitoring of follow-up is largely outside the mandate of OEDE and is driven by the relationship between the Executive Board and Operations, and as such it is reliant on the independence of the Executive Board. The Executive Board has at times been very active in expressing their concerns about weak response, but this may be uneven and it is unclear how their dissatisfaction impacts on programmes and individuals over time.

There are different views as to who should ensure coherence of response to evaluation findings and recommendations, with OEDE feeling it should be ‘management’ and those evaluated often feeling it should be the job of OEDE. Preparation of management response is

⁸ Summary Report on WFP’s Follow-up to Evaluation Recommendations 2005

almost seen as a necessary evil rather than a part of the planning process and is handled in a fragmented manner. Different people produce different parts of the response, which are then merely compiled by one individual who does not represent ‘management’, but at best one division and at worse the rapidly assembled impressions of that individual. The presentations to the Executive Board are sometimes not well prepared. The Executive Director is not always active in defining management response. The result is dissatisfaction by the Executive Board, concerned about WFP’s commitment to utility.

OEDE is currently developing a new follow-up system. OEDE states it will be able to provide WFP management with a status report on the extent to which the recommendations have been implemented. Challenges include (a) the multiplicity of recommendations, (b) that many recommendations are vague, and (c) that many are not seen to be feasible. Some are also concerned that a stricter ‘boxed’ approach will make it even more difficult to deal with the important but unavoidably somewhat more amorphous conceptual/policy-related recommendations. The Panel noted that in future, the OEDE software might well be improved by adding an analytical module and by focusing on improving the nature of recommendations.

5.10 Contribution to building local M&E capacity

In one case (Laos) it was noted that the evaluation process introduced new perspectives and approaches to beneficiary consultation within the Country Office as the external consultant stressed very strongly his convictions that rural people were aware of their own needs and priorities and that they could be trusted. The Country Office reported that this stimulated self-critical reflection on their prevailing approaches which were usually characterised by distrust regarding inappropriate ‘rent-seeking’ behaviour among beneficiaries. This experience was a form of on-the-job awareness raising and capacity building for the staff which was highly appreciated. Some Country Offices see a modicum of capacity building occurring in an on-the-job manner as local staff of WFP and partners accompany evaluation teams. This is confirmed by one NGO partner that reports that their participation in data collection and other activities as part of evaluations is appreciated as a contribution to capacity building.

Despite some positive examples, the objective of building local M&E capacity is clearly not a priority, despite this being an UN Economic and Security Council (ECOSOC) resolution that should apply to WFP.

Some interviewees report that contributions to this goal are more appropriate in relationships with local research institutions for ongoing nutritional monitoring, etc., rather than trying to add this on to already overloaded objectives within three week evaluation missions. Further-

more, Regional Bureaux note that they do not have time to provide a minimum of capacity building for their own staff in evaluation, so it would be inappropriate to even consider trying to use their staff to train others in government or NGOs. On the whole there is a sense among most within the organisation that as an ‘operational organisation’ WFP does not have a strong capacity or responsibility for building local capacity.

This in some ways reflects the invisible nature of partners (especially NGOs), despite the fact that the actual operationalisation of a large proportion of WFP operations is in the hands of partners. It is indeed probably unrealistic to expect WFP field offices to train partners in skills that they lack themselves. The Panel feels that a more appropriate approach to the capacity building issue might be to see how Regional Bureaux and Country Offices can better find ways to share experience with partners in a two-way manner whereby both sides of the partnership learn about how to foster a ‘culture of evaluation’ in their work. More participatory processes would not only improve the quality and use of evaluation results, but also promote learning.

5.11 Conclusions on utility

Overall conclusion

There is insufficient corporate strategic thinking regarding the use of evaluation and where evaluation should fit on the learning-accountability spectrum. Evaluation is rarely used to enhance accountability to partners. The Panel concludes that the criteria of Utility of the WFP evaluation function are partially met with regard to contributing to the formulation of follow-up programmes but that structures and mechanisms to promote utility are weak in most other respects.

Conclusions regarding UNEG norms

Norm	Met?	Comments
Evaluation contributes to managing for results (RBM)	N	Structures for management response at headquarters level are fragmented and weak and require a major overhaul. At decentralised levels evaluation is more directly related to programme formulation and is more effective.
Evaluation contributes to policy making	N	Evaluation is absent from most of the processes by which WFP sets, monitors and analyses policies. WFP rarely uses evaluation as a way of encouraging broader accountabilities in respect to its position within the humanitarian and development communities.
Evaluation checks reality of difference made	N	Evaluation is focused on outputs, as opposed to outcomes and impact. In a narrow sense of 'doing things right' performance is mixed. In a wider perspective of 'doing the right thing' performance is worse. Little consideration is given to alternative, more effective or efficient approaches.
Evaluation is intended for use	P	The intention to use evaluations is evident in respect of the normal timing of programme evaluations and, at least in one case, in its link to real-time policy development.
Evaluation is properly managed	Y	Management of evaluation by OEDE itself is well done, but varies greatly within the broader organisation. At decentralised levels there is a closer link to utility and therefore local solutions are found to use evaluation within a broader learning process, but with a consequent detrimental impact on accountability.
Management responds to evaluation	P	Management responds to evaluation through the process of programme formulation, but in a broader sense this is weak. New plans in OEDE to tie evaluation closely to logframes may enhance utility, but the lack of prevailing understanding and use of logframes will make this path difficult. These problems ultimately stem from the fact that a 'culture of evaluation' and a readiness to acknowledge mistakes is not well embedded into the organisational culture at WFP, nor is it translated into management mechanisms for action.
Evaluation contributes to knowledge building	P	Evaluation makes an insufficient contribution to knowledge building within WFP and virtually none among partners. Access to reports and findings through the website, debriefings, etc., is not sufficiently proactive. Evaluation is not well integrated into other learning initiatives within WFP. Opportunities are missed to distil and disseminate learning systematically. Evaluation does make a major contribution to building knowledge within the Executive Board, particularly through thematic/strategic evaluations and syntheses.
Evaluation contributes to local capacity building	N	Insufficient effort has been made to find opportunities to build local capacities. Many deficiencies noted regarding utility relate to the need to rethink stakeholder engagement throughout the evaluation process.

Y= yes P= partial N = no

6. Principal recommendations

The following recommendations are linked to the findings and conclusions presented earlier in this report. The aim of the recommendations is to suggest ways to improve and amend aspects of WFP's evaluation function where the Peer Panel has identified problems or shortcomings in relation to UNEG's norms and standards or to established evaluation practice.

The recommendations are presented under the headings of a number of issues instead of being sorted along the three main dimensions for the review: independence, credibility and utility. This is intended to make it easier to see to which area or function a recommendation should be applied.

While all recommendations are in principle directed to the management of WFP, i.e. the Executive Director, they are for practical reasons addressed to either WFP or to the evaluation office OEDE, depending on whether it concerns an overall, corporate issue or an issue that can be handled by OEDE directly. As the governing body of WFP they are also directed to the Executive Board for their consideration.

6.1 Relationship between OEDE and the Executive Board

To the Executive Board:

The Panel considers the current reporting lines between OEDE, the Executive Director and the Executive Board as appropriate. However, in order to strengthen OEDE's credibility and perceived independence and to enable deeper engagement and follow-up on evaluations by the Board, the Panel suggest the following measures:

- Establish a Board sub-committee on evaluations in line with existing practice at IFAD. This sub-committee could then be tasked to meet at regular intervals with OEDE to discuss such issues as planned strategies, budgetary allocations, strategic use of evaluations and evaluation follow-up.

- Furthermore, future appointments and contract extensions of the Director of Evaluation should be discussed with the Board prior to their implementation. It would be appropriate to ask a Board Member (possibly the head of the potential evaluation sub-committee) to participate in future interview panels.

6.2 Evaluation policy and strategy

OEDE should develop an evaluation policy that encapsulates and consolidates the previous evaluation policies and fully meets all United Nations Evaluation Group (UNEG) norms and standards for evaluation.

Such an updated and consolidated evaluation policy should i.a. cover:

- explanation of the concept and role of evaluation within WFP;
- explanation of the different types of evaluation applied within WFP as well as between evaluation and other instruments for assessment applied in WFP (e.g. policy research, internal audit, external audit, inspection, after action reviews);
- definition of the roles and the responsibilities regarding the evaluation function within WFP, both at headquarters and at Regional Bureaux and Country Offices;
- the mandate of the OEDE (see also below);
- description of the ways in which evaluations are programmed and prioritised;
- description of the way in which WFP's two-year budget for evaluations as well as the budgets for decentralised evaluations are defined and by whom;
- description of the ways in which individual evaluations are budgeted, organised, managed and adopted;
- mechanisms for the management response of evaluations (both OEDE and decentralised evaluations);
- description of the adherence to WFP's evaluation guidelines and a quality assurance system supporting the decentralised evaluation system.

This policy should be actively shared with the Executive Board, WFP staff, WFP partners and evaluation teams engaged by OEDE or for decentralised evaluations.

The evaluation policy should then be translated into action through the development of an evaluation strategy which would specify how OEDE would implement this policy. The strategy should follow a logical framework format by setting clear goals, outcomes, and outputs, indicate targets and measurable performance indicators. Considering the

current challenges to improve both centralised and decentralised evaluation functions (with a focused on improving learning and accountability), it may well be that the strategy will imply increasing the funding for the evaluation function in WFP. Given the fact that this is unlikely to be a realistic proposition, the Panel nevertheless feels that in the medium-term OEDE should approach key donors for short-term funding of specialist staff positions, junior professional officers, evaluation posts in Regional Bureaux and support to modify, develop and test appropriate tools and guidelines.

The role and purpose of and the relationship between (a) self-evaluation, which essentially means self-reflection on performance using, for example, AARs as one method, (b) decentralised evaluations (where Country Offices commission evaluations of their own interventions using external evaluators), and (c) external evaluation (where OEDE, which has no involvement in programs, commissions evaluations) should be studied. It should be clearly articulated in both the evaluation policy and also in the overarching policy and strategy documents of WFP.

6.3 OEDE's mandate

The mandate for OEDE should form part of WFP's evaluation policy. The OEDE mandate should include:

- undertake (a) strategic (thematic) evaluations of WFP's management and programme policies, conduct evaluations of specific policies in the organisation and (b) evaluate the outcomes and impact of WFP funded programmes and projects (both humanitarian and development);
- promote the use of evaluation findings, recommendations and lessons in policy development and programme/project formulation;
- ensure that evaluation findings are accessible throughout the organisation and to WFP's stakeholders (Executive Board, cooperating partners and other stakeholders) and that evaluation results and lessons are build into WFP's management information system;
- develop and disseminate evaluation guidelines, methods and tools to assist OEDE staff and programme managers in the conduct of evaluations;
- develop methodological tools and systems to support results orientation in WFP (the OEDE should establish a closer working relation with Oversight, the Change Management Unit and units currently responsible for 'mainstreamed' RBM);
- engage in continuous interaction with Country Offices and Regional Bureaux to foster the internalisation and application of evaluation standards within WFP as well as among cooperating partners;

- oversee the evaluation function of WFP, including compliance with and tracking of recommendations adopted in the management response to evaluations;
- prepare the Annual Evaluation Report which is submitted to the Board (preferably the Annual Evaluation Report should include a section on how the lessons and recommendations have been followed up as well as a section on the quality of decentralised evaluations);
- maintain and develop partnerships within the United Nations Evaluation Group (UNEG), ALNAP, and other appropriate international and regional evaluation networks in order further develop the quality of evaluation and the harmonisation of evaluation practices.

6.4 External accountability; relations with partners and stakeholders

To OEDE:

A key challenge for WFP's accountability is that, while institutional accountability tends to be 'upwards' towards their Board and major donors, WFP's accountability towards intended beneficiaries is mainly dependent on the interface between the host government and/or NGO cooperating partners.

Recommendations

- OEDE should develop an 'accountability map' of key WFP stakeholders, both internal and external, to help in clarifying roles and responsibilities.
- Based on this 'accountability map' develop guidance both for WFP staff in functional units and partners identified in the map as key stakeholders to help them in fulfilling their accountability responsibilities and enhancing communications.

An example of a stakeholder "map" linked to a communication strategy is shown below. While this implies that WFP staff may need to adopt different approaches with certain stakeholder groups and reprioritize resources, recommendations below attempt to illustrate how participatory approaches can help spread the workload and make more effective use of limited resources.

Key stakeholder groups	What info do they need?	What do they need info for?	When will they need it?	How should it be communicated?*	Who should communicate it?
WFP field staff					
Host govt					
NGO Partners					
Communities					
Donors					
WFP Board					

* Typical examples are workshops, stakeholder meetings, list(s) of FAQs, summary docs, talking points, bilateral meetings, conferences, role-play, video, policy papers, guidelines to WFP.

Recommendation

WFP should look for ways of promoting, and providing incentives for staff to adopt more participatory approaches in evaluations. While engagement with partners (whether at a country, regional or global level) is normally with WFP staff outside OEDE, OEDE could play a facilitatory role in terms of:

- emphasising participatory practices in a revised evaluation policy;
- adapting, adopting and disseminating relevant guidelines; OEDE staff involved in evaluations, either as a manager and/or team member, can model or facilitate participatory processes.

A genuine participatory approach requires substantive involvement of relevant stakeholders at each stage of the programme and project cycle, not only at the evaluation stage, and there is extensive ‘good practice’ guidance available. WFP staff should thus already be using participatory approaches with partners anyway and the responsibility of the OEDE evaluation manager and evaluation team leader is to make optimal use of participatory systems already in place during the planning, implementation and utilisation of the evaluation.

As described in the report, findings during the current Peer Review suggest that participatory approaches are not widely applied which implies that OEDE has a role to play in building capacity of WFP field staff to:

- Ensure that WFP field staff are provided with appropriate support and guidelines to facilitate participatory approaches during evaluation processes. This may start with guidance to ensure that WFP staff are aware that sharing of draft ToRs, reports, etc. with external stakeholders is not only authorized, but also encouraged.
- OEDE staff should as much as possible “model” participatory approaches, both in their roles as an evaluation manager or, periodically, as a team member. OEDE staff could also assist in facilitating or

advising on Country Office-organised workshops to disseminate results of evaluations.

- OEDE should use communication and learning strategies to support the above efforts.

6.5 Management response

The evaluation policy of 2003, where the responsibility for management response is delegated to the Director of the now defunct OEDR⁹ under whom the OEDE was located, may be one reason why the lines of responsibility are blurred regarding this issue. However, it seems that the OEDE for some time has contributed to the lack of clarity by accepting, at least in some cases, to coordinate the response before an evaluation is submitted to the Executive Board.

- WFP should, both in principle and in practice, establish a clear division of responsibility regarding management response between the evaluation function and the organisation's line management. After an evaluation has been submitted to the Executive Director the Evaluation Office should not be involved with drafting or compilation of responses from different parts of the organisation; the general principle is that the Executive Director has the overall responsibility for management response whether the actual drafting is delegated or not to other parts of WFP.
- The management response mechanism should include rules about the timeframe for the response and procedures for follow-up of the management response as well as for reporting to the Executive Board and the OEDE about the results of the follow-up.
- The management response should, whenever appropriate, distinguish between short term and long term responses as well as between operational measures directly related to the subject matter for the evaluation and general lessons to be learnt by WFP and its partners.
- The management responses should include justification for not accepting a specific recommendation.
- A similar system for management response should be used for decentralised evaluations. The same kind of division of responsibilities can for obvious reasons not be established when e.g. a Country Director both commissions an evaluation and decides on management response but it is essential that ways are created for formal re-

⁹ The 2003 evaluation policy says in paragraph 30: "...the Director of OEDE will also formally submit the summary reports to the Executive Director through the Director of OEDR, thus permitting the Executive Director to prepare management's response to the recommendations for the Executive Board as an information paper. Responsibility for management's response will be delegated to the Director of OEDR, whose final responsibility will be to ensure that the lessons learned from monitoring and evaluation are translated into action."

sponses to such evaluations. For example, when an evaluation is carried out by external evaluators a Country Director can still independently formulate a management response and report to the Regional Bureau Director or the Executive Director. Follow-up should also be the Country Office's responsibility with reporting on the results upwards in the organisation.

- The management response and follow-up mechanism should be transparent with relevant documents easily accessible for WFP and partners and routinely posted in electronic form.
- Mechanisms should be found to improve the quality, credibility and ownership of evaluation recommendations which form the basis for management response. Such mechanisms may include developing recommendations in dialogue with primary stakeholders, and/or leaving recommendations up to those responsible for decisions and action in the organisation, based upon engagement by primary stakeholders around the findings and conclusions of the evaluation report.

6.6 Evaluation quality

The current emphasis placed on improving the quality, rigor and harmonisation of OEDE's work as well as the focus on systematic processes, quality checks and tools such as stakeholder maps and evaluation matrices are highly encouraged. In addition to the ongoing systematisation of processes and the development of templates and codes of conduct the Panel recommend that:

- all older documents that might cause confusion be recalled;
- Terms of Reference of evaluation teams include the stipulation that they should adhere as far as possible to the chosen professional evaluation standards or guidelines (for example the UNEG Norms and Standards; African Evaluation Guidelines; International Program Evaluation Standards). Justification should be provided in the report for key standards that were inappropriate or not met;
- further institutionalise the practice of scoping and inception missions to be undertaken by OEDE evaluation managers and team leaders to (a) help determine evaluability; (b) focus the evaluation; (c) increase the buy-in among, and contribution by primary stakeholders; and (d) help shape the final Terms of Reference and form the basis for the inception report;
- ensure all reports contain a comprehensive description of the methodology, including but not limited to the rationale for the choice of methodology, the final evaluation matrix, ethical considerations and technical constraints and level of adherence to appropriate evaluation standards;

- meta-evaluations should be done – as intended by OEDE – on all strategic and selected decentralised evaluations. Care should however be taken that they are conducted by evaluation experts who are well acquainted with standard meta-evaluation frameworks even if tailor-made frameworks are used for WFP.

The capacity of OEDE staff should be maintained over time to stimulate interest in the evaluation field and encourage professionalism. We recommend that:

- ample time should be allocated and incentives should be provided for staff to keep up with new developments in the field of evaluation.

6.7 Organisational learning

The tension and complementarities between evaluation for accountability and for learning seem not to be acknowledged everywhere in WFP. This requires more attention. Ideally the learning element should be linked to a larger organisational knowledge management strategy.

Recommendations

OEDE should establish mechanisms to systematically harvest lessons from existing evaluations. Such lessons should then be proactively shared, using internal knowledge management fora and tools such as the PASSiton, as well as external knowledge sharing fora such as ALNAP, the IASC and relevant partners.

Innovative methods for extracting and sharing of evaluation lessons should be investigated, building on the experiences of other organisations with extensive experience in this field. These would include

- creating greater ownership of evaluation results through the manner in which the evaluation process is managed;
- investigating the potential of various tailor-made communication methods for reporting tentative and final evaluation results;
- conducting frequent meta-analyses across evaluations on themes of importance to WFP. This presupposes that the evaluations will be designed in a manner that will enable extraction of contextualized lessons on such themes;
- identifying organisational disincentives for learning and knowledge and addressing them.

6.8 Monitoring and RBM

To WFP:

WFP should give high priority to address the disconnect between its various results-focused data collection, reporting and analysis tools. A thorough review of existing field monitoring systems and applications is vital to ensure that evaluations as well as the corporate monitoring systems have access to more reliable, relevant and comparable data.

Ways should be developed and maintained to ensure that all interventions are linked to proper monitoring mechanisms, both at local and corporate levels, and include objectives and indicators that facilitate evaluations which satisfy WFP's as well as external stakeholders' needs.

6.9 Team selection and procurement of external evaluation expertise

To OEDE:

It is recommended that OEDE develops a transparent, rigorous and competitive approach to the selection of team leaders. This should include advertising the evaluation consultancies on appropriate listserves, shortlisting based on expression of interest, shortlisting and selecting team leaders based on the submission of an approach note and on interviews. If possible, team leaders should be identified early on and be involved in the identification and selection of the rest of the team.

All evaluation teams should include at least one evaluation specialist, preferably the team leader, who has sufficient knowledge about and experience with current evaluation approaches and methods.

6.10 Staffing of OEDE

The peer review considers the current mix of internal WFP career staff and externally recruited professional evaluators suitable. However, given the technical nature and professional skills profile of these positions, the Panel recommend WFP to

- allow OEDE to select internal staff based on a professional recruitment process rather than through the standard reassignment exercise. This should include a selection process based on the staff members' interest in work in OEDE, the extent to which their competencies match the ones needed by OEDE, and a competency-based interview of the top three candidates;
- continue to allow external recruitment of evaluation specialists;
- base OEDE's staff profile on the profile of evaluators developed by UNEG;
- consider how to ensure an appropriate career path for evaluation specialists within the organisation and within the UN system.

6.11 Budget for evaluations

To address concerns that only a small portion of the overall evaluation budget is within the direct control of OEDE, WFP's senior management should devise ways to safeguard the considerable funding allocated to evaluations for the next biennium. The use of Direct Support Cost (DSC) of projects appears a budgetary necessity for the time being but it is not an ideal situation. Based on prior experience with this model, it is critical that the Executive Director and senior management ensure the full use of these funds by holding managers accountable for (not) implementing decentralised evaluations. Furthermore, management should consider to ' earmark ' strategic and sensitive decentralised evaluations for OEDE-management to thus reduce the risk that offices seek to bypass OEDE and to ensure full independence where most needed.

The establishment of a centrally managed fund (budget line) for evaluation (both OEDE evaluations and decentralised evaluations) should be investigated.

Appendix 1

Approach and Work Plan

Introduction

The Professional Peer Review is conducted in line with the *Framework for Professional Peer Reviews of Evaluation Functions in Multilateral Organizations*, which was finalized by the DAC/UNEG Joint Task Force in early 2007. Following the peer reviews of the evaluation function of the UNDP and UNICEF, the World Food Programme (WFP) volunteered as the third multilateral organization for such review.

Experiences from the previous two Peer Reviews have been taken into account in this document which sets out the key elements of the Peer Review of the evaluation function of the WFP. It describes the background of the Peer Review, its purpose, the scope and general approach and methods, the composition of the Peer Panel and the time schedule. The document, which was agreed by the Panel members and has been shared with WFP for comments, serves as a basic reference guide for the Review.

Background

WFP formally requested the Chair of the DAC/UNEG Task Force to initiate the assembly of the Panel on 20 December 2006. The Chair approached Sida's Department for Evaluation and Internal Audit to take the lead in this Peer Review, which would involve making arrangements with the Office of Evaluation at WFP to plan the Review, establishing the Peer Panel and recruiting consultants to serve as advisors to the Panel. This was accepted by Sida at the end of January 2007.

In a contribution to this Approach Paper the Office of Evaluation at WFP describes the development of its evaluation function in the following way:

“WFP has had an evaluation function since 1965. The Office of Evaluation (OEDE), and its predecessors, has undergone a number of changes since. Between 2000 and 2005, the Executive Board (EB) of WFP received four papers on the evaluation function of WFP¹⁰. The main concerns of the EB members were (i) location, (ii) reporting lines, (iii) budget, and (iv) interface with the Board.

During the Board’s 2006 annual session, the Executive Director of WFP agreed to professionalizing OEDE by (a) appointing a director with a proven track record in evaluation (rather than filling the position through an internal appointment), (b) upgrading the position and making the director of evaluation part of the executive staff, (c) changing the reporting lines, i.e. the director reports now directly to the Executive Director.

At the same time, WFP informed the member countries of WFP about the intention to volunteer for a professional peer review of the evaluation function. As a result, the EB members agreed to hold further queries and discussions about the independence of OEDE until the results of the Professional Peer Review were presented to the EB.”

Purpose of the Professional Peer Review

The purpose of the Professional Peer Review is to provide DAC and UNEG members as well as decision-makers in the leadership of WFP, WFP’s Executive Board members and the OEDE with an independent assessment of the functioning of OEDE and the quality of its work. The findings of the Professional Peer Review will be presented to WFP’s Executive Board meeting in February 2008 and inform further discussions and decisions about the functional and administrative independence of OEDE.

The Professional Peer Review takes the central evaluation function, i.e. OEDE as its starting point but will include also the decentralised evaluation work in the review. It reviews the evaluation function in light of the objectives and structure of the WFP and according to the core assessment questions summarised below.

¹⁰ WFP’s Principles and Methods of Monitoring and Evaluation, presented for consideration at WFP/EB.A/2000/4-C; A Policy for Results-Oriented Monitoring and Evaluation in the World Food Programme, presented for approval at WFP/EB.A/2002/5-C; WFP’s Evaluation Policy, presented for approval at WFP/EB.3/2003/4-C; and Report on the Management of Evaluation, presented for consideration at WFP/EB.A/2005/5-E. The four documents can be accessed at www.wfp.org/operations/evaluation.

Scope of and Limitations to the Professional Peer Review

The Professional Peer Review will examine and comment on:

8. Structural aspects of how the evaluation function operates in WFP, including whether the current functional arrangements are effective in ensuring that OEDE can contribute to the learning and accountability within WFP;
9. The evaluation policy of WFP and other policies and procedures having a bearing on OEDE and its work, in particular the extent to which the evaluation policy conforms with international standards, and whether other policies are relevant to the functioning of OEDE (e.g. those concerning results-based management, strategic planning, budgeting, evaluation coverage of operations, decentralized evaluations, etc.).
10. Organizational relationships of OEDE at the governance level (WFP's Executive Board and Bureau), OEDE's position and relationship to the Executive Director, the Core Management Team, and the Executive Staff, the roles and responsibilities of OEDE vis-à-vis the Regional Bureaux and the Country Offices concerning decentralized evaluations; the roles and responsibilities of OEDE in relation to other HQ departments (e.g. Policy, Operations etc).
11. Relationships regarding the evaluation function and responsibilities vis-à-vis WFP's cooperating partners.
12. The quality of the evaluations undertaken and commissioned by OEDE and by regional and country offices. This includes the conduct of the actual evaluation, the quality of the evaluation reports, the independence of evaluation teams and team leaders (consultants), the ways in which OEDE enables them to produce credible reports including the ways stakeholders are facilitated to comment on draft reports (e.g. when do comments become an infringement on independence and when are they warranted to ensure standards of evaluation reports).
13. Use of evaluation results and follow-up. Important aspects are: the ways in which evaluation results are disseminated and lessons used both within WFP and by others (donors, cooperating partners etc); the responsibility for the follow-up of recommendations with management; and how follow-up is undertaken and monitored.

The Peer Review will collect and analyse information about evaluations managed both by WFP HQ and field offices (regional and country level) and include a discussion of decentralised approaches towards evaluations. This discussion will consider the oversight function of OEDE in relation to evaluations led by WFP regional or country offices.

Core assessment questions

In line with the *Framework for Professional Peer Reviews of Evaluation Functions in Multilateral Organizations*, the Peer Review of the Evaluation function of WFP will apply three core criteria that need to be satisfied for evaluation functions and products to be considered of high quality:

- G. *Independence* of evaluations and the evaluation system(s). The evaluation process should be impartial and independent in its function from the process concerned with the policy making, the delivery, and the management of assistance. A requisite measure of independence of the evaluation function is a recognised pre-condition for credibility, validity and usefulness. At the same time, the review should bear in mind in that the appropriate guarantees of the necessary independence WFP is defined according to the nature of its work, its governance and decision-making arrangements, and other factors. Moreover, like most organizations WFP's aim is to encourage the active application and use of evaluations at all levels of management, meaning that systemic measures for ensuring the necessary objectivity and impartiality of this work should receive due attention.
- H. *Credibility* of evaluations. The credibility of evaluation depends on the expertise and independence of the evaluators and the degree of transparency of the evaluation process. Credibility requires that evaluations should report successes as well as failures. Recipient countries should, as a rule, fully participate in evaluation in order to promote credibility and commitment. Whether and how the organization's approach to evaluation fosters partnership and helps builds ownership and capacity in developing countries merits attention as a major theme.
- I. *Utility* of evaluations. To have an impact on decision-making, evaluation findings must be perceived as relevant and useful and be presented in a clear and concise way. They should fully reflect the different interests and needs of the many parties involved in development co-operation. Importantly, each review should bear in mind that ensuring the utility of evaluations is only partly under the control of evaluators. It is also critically a function of the interest of managers, and member countries through their participation on governing bodies, in commissioning, receiving and using evaluations.

The advisor(s) to the Peer Panel will together with the Peer Panel prepare a detailed set of assessment questions related to each of the core questions in order to better focus the review. This set of questions will be formulated taking into account similar questions in the previous peer reviews.

Normative framework

WFP's evaluation office is part of two professional evaluation groups: that of evaluators of humanitarian assistance (ALNAP), which includes *inter alia* bilateral and multilateral organizations and international NGOs, and of UNEG, composed of 36 heads of evaluation functions of the UN System and works in close cooperating with the OECD/DAC Evaluation Network. The Peer Review Panel will consider the normative framework of UNEG, and relevant guidelines of the OECD/DAC Evaluation Network as well as ALNAP's proforma for evaluation quality assessment of evaluation reports when assessing WFP's evaluation function. This will include efforts to harmonise the various guidelines in order to facilitate the work of the Peer Panel.

Panel composition

A number of important considerations were taken into account when composing the panel membership: (i) relevant professional experience – WFP has a dual role in providing humanitarian and development assistance and therefore it was important that panel members brought together professional experience of both types of work; (ii) independence – to avoid any potential or alleged conflict of interest or partiality, the panel members should not have any close working relationship to WFP that might influence the Panel's position and deliberations; and (iii) broader membership – experience and viewpoints from donors, executing organisations and partner countries should be represented in the panel.

The combination of these criteria together with the voluntary nature of serving on the Panel resulted in the following composition:

Jock Baker, accountability & programme quality coordinator, CARE International

Stefan Dahlgren, senior evaluation officer, Sida (lead responsibility)

Susanne Frueh, former chief of evaluation, OCHA

Ted Kliest, senior evaluation officer, Netherlands Ministry for Foreign Affairs

Zenda Ofir, independent member

The Panel will be assisted by two advisors responsible for data collection and information gathering; preliminary assessment the collected information which is to form the basis for more detailed information gathering through structured and semi-structured interviews. The advisors will provide the Panel with a consolidated information base, specifying the sources. With the benefit of the information assembled by the advisors, its examination by the members of the Peer Panel, and observations provided by WFP on the information gathered, the Peer Panel will conduct interviews with WFP Senior Managers, other senior staff and (a selection of) Executive Board Members.

Reporting

The Peer panel will submit its report to WFP's OEDE, which in turn will provide it to the Executive Director of WFP, the Executive Board and disseminate it within WFP. The final report will also be provided to the DAC/UNEG Task Force, for dissemination among its respective constituencies and to interested cooperating partners. The Peer Panel will from time to time report on the Review's progress to OEDE and the DAC/UNEG Task Force.

Responsibility of OEDE

OEDE serves as the main contact point within WFP for the Panel and its advisors. OEDE will provide requested information and data, including the names and details of contact persons whom the Panel or its advisors wish to contact.

OEDE will continue to brief WFP Executive Staff (through the Internal Evaluation Committee) and the Executive Board (through an Informal Meeting) about the Peer Review. OEDE will also be responsible for submitting the Panel's report and recommendations to the Executive Board and for reporting on follow-up action.

OEDE aims to provide the DAC/UNEG Task Force with feedback on the experience of the Panel Review to enable the members of the DAC Evaluation Network and the members of UNEG to learn from WFP's experience.

Review process

The Peer Review will employ the following steps:

1. Preparation of the Approach and Work Plan and Terms of Reference for the advisors to the Peer panel.
2. Initial meeting of the Peer Panel to discuss details of the task as outlined in the Approach and Work Plan and ToR for the advisors and to familiarise itself with WFP's evaluation work.
3. Preparatory work including a desk review and interviews undertaken by the advisors. During this phase the advisors will analyze relevant documentation and carry out interviews at WFP headquarters in Rome, a selection of WFP's regional and country field offices (preferably by telephone), former evaluation team leaders and with representatives of member countries (in particular Executive Board members). Stakeholders include: Executive Board members, the Executive Director, Executive Staff (individually and as members of the Internal Evaluation Committee), OEDE (director and staff), regional and country directors, division directors, and staff. In addition views from external stakeholders like cooperating partners will be taken into account to the extent possible. Part of the data collection from field offices will be carried out through an e-mail survey.

4. Analysis and identification of issues for in-depth discussion – the desk review and preliminary interviews will generate issues for follow-up by the Panel members. The advisors will present a draft factual report to the Panel.
5. Peer Panel interviews with selected stakeholders based on the issues raised by the advisors. Preliminary assessment by the panel, which will include (a) Panel agreement on its framework for judgement, (b) Panel consideration of evidence and findings in order to arrive at draft conclusions and recommendations and (c) agreement on an outline for the draft report. The preliminary assessment report will be drafted by the advisors.
6. Peer Panel finalises the draft assessment report.
7. Draft assessment report is discussed with WFP in a Review Meeting. Based on this discussion the final assessment report is produced.
8. The final assessment report will be submitted to WFP (the report will be submitted by OEDE to the WFP Executive Board's 2008 first regular session in February 2008). The final assessment report will also be provided for information to the DAC/UNEG Task Force.

Schedule

The Peer Review will be conducted according to the following preliminary schedule:

Composition of the Panel	January–March 2007
Approach and work plan	End of March 2007
Recruitment of Advisors	March–April 2007
Initial meeting of the Peer Panel	April 2007
Preparatory Work by Advisors	May–July 2007
1st report – to the Panel meeting in Geneva	25 June
2nd report to the Panel	August 2007
Panel Visit to Rome (interviews)	End of September 2007
Finalization of Draft Report	2nd Week October 2007
Review Meeting on Draft Final Report in Rome	25–26 October 2007
Final Report	End October 2007
Discussion of Draft Final Report in WFP's Executive Board	February 2008

Throughout the period, consultations will take place between the Panel lead agency and OEDE, within the Panel, and between the Panel and the advisors, as necessary.

The need for Regional Bureaux or country office visits will be determined during the process of the Peer Review, keeping in mind the option to conduct telephone interviews.

Resources

The costs of the respective Panel members should be covered by her or his agency; the costs of the independent member of the Panel will be covered by Sida acting as the lead agency for the Review. The cost of the advisors will primarily be covered by Sida; other DAC or UNEG members may provide financial contribution on a voluntary basis.

OEDE's contribution to the exercise will be in-kind (professional and general service staff time for organizing and facilitating the Peer Review process).

Appendix 2

Framework for Professional Peer Reviews

By DAC/UNEG Joint Task Force on Professional Peer Reviews of Evaluation Functions in Multilateral Organizations 8 January, 2007

Summary

This framework for professional peer reviews of evaluation in multilateral organizations is one contribution to the efforts in the international community to strengthen a shared agenda for improving performance in development and cooperation. Since 2004, the community of evaluators in international cooperation agencies have themselves developed and tested (through pilot peer reviews in UNDP and UNICEF) a new approach to assessing and enhancing multilateral agencies' own evaluation capacity and performance. Now the DAC Evaluation Network and the UN Evaluation Group have joined forces to establish an internationally recognised form of peer review of the evaluation function. The joint task force will focus on developing a framework for professional peer reviews, based on previous experiences and internationally recognized standards.

This approach has several purposes: building greater knowledge, confidence and use of evaluation systems by management, governing bodies and others; providing a suitable way of "evaluating the evaluators"; sharing good practice, experience and mutual learning. The primary intended audience for the results of these professional peer reviews is one of decision-makers and other users of evaluation – including where appropriate the intended beneficiaries in member countries.

On the basis of substantial analysis and lessons from the pilot experiences, this framework sets out both specific recommendations and some options for the *"Approach and Methodology"* for these professional peer reviews, beginning with an explanation of the nature, strengths

and limits of a peer review. Because the peer reviews are intended to assess the evaluation function against accepted international standards in a wide range of organizations, the framework recommended here has a blend of standard and flexible elements. It sets out the specific focus of the peer reviews of evaluation.

The framework contains clear guidelines on a consistent, broad normative framework using applicable norms and standards, and also provides some tested, step-by-step options and practical suggestions for carrying out the review. At the same time, it is intended to allow for necessary flexibility, for example to adapt for de-centralised and centralized evaluation systems or to feature special issues or emphases for particular organizations. It is worth noting that while the framework has been primarily designed for reviews of multilateral development agencies, it should prove equally useful (with limited adaptation) for any evaluation function such as those of international financial institutions, individual bilateral donor agencies, or other multilateral agencies, for example those concerned with humanitarian relief or human rights issues.

The framework points out there are a number of issues that are likely to require decision as the plan is finalized for each individual review exercise. It is also clearly recognized that work may need to be done to adapt the model much further in some cases – for example, to very small or highly specialized organizations, and/or those with extremely limited existing evaluation capacities. The framework also includes an explicit treatment of the risks that may arise in these reviews, and ways of managing them.

The Framework for Assessment is outlined, including a core assessment question. As reflected in this question, the approach and methodology hinges on using recognized normative frameworks and clustering the treatment of the many issues under three crucial criteria for evaluation, specified as:

1. “The independence of evaluations and evaluation systems” – Noting that this attribute is never absolute, and needs careful analysis in different contexts;
2. “The credibility of evaluations”. This includes assessment of whether and how the organization’s approach to evaluation fosters partnership and helps build ownership and capacity in member countries where appropriate; and
3. “The utility of evaluations” – Recognizing that this is only partly under the control of evaluators, and is also critically a function of the interest of managers, and member countries through their participation on governing bodies, in commissioning, receiving and using evaluations.

Depending on their applicability to the organization being reviewed, one of several sets of accepted and consistent standards should be applied to provide a normative framework for the assessment under these three broad headings. To avoid future professional peer reviews of UN organizations having to repeat the necessary sorting and categorization work needlessly, such review teams are referred to the documentation of the UNDP and UNICEF peer reviews to be able to extract readily the relevant materials, and similar work was done for the UN Norms and Standards and the DAC Principles. These peer reviews may be one of the most important continuing general tests of the usefulness of the prevailing norms and standards for evaluation, and they should thus be prepared to suggest ways to strengthen the norms on the basis of these experiences.

The experience of the pilot reviews suggests that it may normally be appropriate to apply a combination of indirect (evaluation system) and direct (sample evaluation product) assessments at different levels. Although there are not yet single, internationally accepted standards for assessing the quality of evaluation products, a number of useful models are already being tested and applied at different levels.

Given the central importance of the Normative Framework to the whole peer review process and its results, experience confirms the importance of the specific framework being agreed between the peer review team and organization at an early stage.

The framework outlines the key roles and responsibilities of different parties in the professional peer reviews, beginning with the peer Panel, including appropriate criteria for selection of its members, and the main expectations for its tasks. Similarly, it outlines the major roles and responsibilities that will fall on the organization under review, which will normally be carried mainly, but not exclusively, by its evaluation unit.

To facilitate the work of those launching future reviews, the main lines of the approach, and some important practical lessons, are illustrated with typical steps and sequences, in a section providing a “phase-by-phase organization of a peer review process”. The Framework document concludes with brief discussions of the time and resources that are likely to be required for a professional peer review of evaluation, and of the planned monitoring and assessment of experience in these reviews.

Background

In the field of international cooperation, the international community has come together around a shared agenda for improving performance in development and cooperation. As part of this drive, the need for strong monitoring and evaluation has been a key concern. It is closely linked to setting clearer objectives, improving working practices, and applying the lessons of experience – all with member countries and

their people more firmly in the “driver’s seat”, where they must be if durable progress is to be achieved.

All the key actors in international cooperation are now part of this drive for evidence of effectiveness and improvement, including the multilateral development agencies that channel a large share of total assistance flows to developing, transitional and other member countries. The member countries of these agencies need credible evidence of effectiveness and benchmarks for improvement, as do their partners on the ground, as do their own senior managers and operational staff. At the same time, assessing the effectiveness of these organisations can pose particular challenges because of their wide membership, international governance and sometimes the range and types of their activities.

After the experience in recent years of several major external evaluations of multilateral agencies – mostly organised by consortia of funding governments – two communities of evaluators in development agencies¹¹ have since 2004 themselves developed and tested a new approach to assessing and enhancing multilateral agencies’ own evaluation capacity and performance. It is worth noting that while the framework has been primarily designed for reviews of multilateral development agencies, it should prove equally useful (with limited adaptation) for any evaluation function such as those of international financial institutions, individual bilateral donor agencies, other multilateral agencies, for example those concerned with humanitarian, relief or human rights issues, as well as global programs, foundations and global non-governmental organizations.

Rationale

This approach has several purposes:

1. Building greater knowledge and use of evaluation systems by management, governing bodies and others, leading to: better understanding of the current quality and needed improvements in evaluation, better integration of evaluation priorities and findings into performance management, improved evaluation policy and practice at all levels, and stronger planning and budgeting for evaluation;
2. Within an organization, providing a way of “evaluating the evaluators” through a professional peer assessment against recognized international standards, thus respecting the necessary degree of independence of the evaluation function from direct assessment by management;
3. Sharing good practice, experience and mutual learning, building greater internal capacity and external confidence in multilateral evaluation systems and thus ultimately reducing demands for special outside assessments of performance.

¹¹ The OECD/DAC Network on Development Evaluation, and the UN Evaluation Group (UNEG).

On the invitation of the Evaluation Offices of UNDP and UNICEF, pilot peer reviews of evaluation in those two organizations were carried out and published in 2005–2006, and the experiences of both exercises were documented and assessed by the participants.¹² Both the pilot reviews were found to have been extremely valuable, in distinct ways, for the two organizations and panels concerned.

The DAC Evaluation Network and the UN Evaluation Group have now joined forces to establish an internationally recognised form of professional peer review of the evaluation function. The joint task force has focused on developing this framework for professional peer reviews, based on previous experiences and internationally recognized standards. It will promote peer reviews, starting with interested UN organizations. The joint task force sees its work as catalytic in ensuring that peer reviews will incorporate perspectives from a broad range of professional expertise in (inter)national organizations, including perspectives of member countries, and non-governmental and private sector expertise.

Audience. In view of the purposes and experience described above, the primary intended audience for the results of these professional peer reviews is one of decision-makers and other users of evaluation – including where applicable the intended beneficiaries in member countries – although there is much to be gained from the exchange between evaluation specialists. Given this primary audience, it is especially important to avoid engaging or appearing to engage in special pleading or professional courtesy among professional evaluators, and to ensure that the review embodies objective and demanding standards for advancing organizational performance and accountability.

Approach and methodology: Recommendations and options

The overall approach

The nature, strengths and limits of a professional peer review.

Peer review can be described as the systematic examination and assessment of the performance of an organization by its peers, with the ultimate goal of helping the reviewed organization improve its policy making, adopt best practices, and comply with established standards and principles. The examination is conducted on a non-adversarial basis, and it relies heavily on mutual trust among the organizations involved in the review, as well as their shared confidence in the process.

It is not intended to serve as a procedure for resolving differences and peer review never implies a punitive decision or sanctions; it generally goes beyond fact-finding to include an assessment of the performance, and is characterized by dialogue and interactive investigation. Peer pressure does not take the form of legally binding acts, as sanctions or other enforcement mechanisms. Instead, it is a means of peer persua-

¹² See references 2 and 3

sion which can become an important driving force to stimulate organizations to change, achieve goals and meet standards.¹³

All professional peer reviews contain the following structural elements, and the following sections of this framework outline recommendations and options for applying them to these reviews of the evaluation function in multilateral organizations:

- A basis for proceeding: including the necessary agreement for cooperation, and a clear question for assessment;
- A normative framework: an agreed set of principles, standards and criteria against which performance is to be reviewed;
- Designated actors and roles in carrying out the peer review; and
- A set of procedures for planning, assembling and testing the base of evidence and findings, leading to the final result of the peer review – an exchange around the conclusions drawn primarily from the diverse professional experience and judgments of the peer panelists themselves, deliberating as a group.

The device of professional peer assessment, already well tested in the development field, has been agreed to bring a number of special strengths to this undertaking, as illustrated in international processes such as those of the OECD, WTO and now NEPAD. First, it starts with a shared appreciation of the distinctive challenges of work and evaluation in international cooperation, and the fact that all concerned are constantly striving to improve. Second, it can adapt and apply the most pertinent professional principles, norms and standards in coming to an assessment. Third, drawing on experienced professional peers from other institutions (participating as individuals) will maximize the opportunities for sharing relevant experience and lessons. Finally, the resulting assessment should carry particular weight, both internally and externally, for the independence and professional credibility of its results. All these aspirations highlight the importance of the quality of the panel and its support, as well as the openness and cooperation of the organization being reviewed.

To avoid unrealistic expectations, it is important to be clear that a professional peer review of evaluation in an organization is not in itself an assessment of the effectiveness of that organization. However, it can and should contribute to the basis for assessing the effectiveness of the organization, by testing the capacity and quality of the organization's own evaluations of effectiveness, and thus the confidence that can be placed in them. Further, while a peer review of the evaluation function is focused on one important part of an organization's performance

¹³ The basic parameters of peer review outlined here are taken from "Peer review: a tool for cooperation and change: An analysis of an OECD working method," by Fabrizio Pagani. Organisation for Economic Co-operation and Development, Paris, 2002. They have been adapted to the special character of reviews under this framework.

management systems, at the same time it is also likely to shed light on other parts, such as the framework for managing for results, etc. Maintaining this specialised focus on evaluation will be important for the reviews, while also taking into account how it actually works with these inter-related systems in the organization.

It should also be stressed that these peer reviews:

- Are not themselves full-fledged evaluations – they are less comprehensive and in-depth assessments but, as outlined below, they should adhere to a rigorous methodology – applying the key principles of evaluation while taking full advantage of the particular benefits of a peer mechanism. It is explicit that the final conclusions will clearly be a judgment by the Panel concerned.
- Must themselves reflect the accepted standards of good practice in development cooperation by seriously engaging developing country partners in the process of the review, in the learning acquired, and in the use of the results.
- Are designed to be targeted and lean – without sacrificing the rigour required for their basic validity and credibility – to avoid making unreasonable demands of time, expense and additional workload on both Panel members and the organization being reviewed.

Because the professional peer reviews are intended to assess the evaluation function against accepted international standards in a wide range of organizations, the framework recommended here has a blend of standard and flexible elements. Thus it contains clear guidelines on a consistent broad normative framework using applicable norms and standards, and also provides some tested, step-by-step options and practical suggestions for carrying out the review. At the same time, it is intended to allow for necessary flexibility for example, to adapt for decentralised and centralized evaluation systems or to feature special issues or emphases for particular organizations. Thus there are a number of issues that are likely to require decision as the plan is finalized for each individual review exercise. It is also clearly recognized that work may need to be done to adapt the model much further in some cases – for example, to very small or highly specialized organizations, and/or those with extremely limited existing evaluation capacities.

Finally, because the approach is new and still being refined it is not intended, to be overly prescriptive. Experimentation is to be expected, and the documentation and sharing of experience and lessons is encouraged, with a view to possible further formalization (e.g. rotational coverage vs. volunteering) on the basis of the experience of a substantial number of further reviews.

Risks and their management

Among the possible risks that may be encountered in carrying out these professional peer reviews, four appear to be most prominent:

- i. The review might become too heavy to be justified – e.g. very burdensome on the organization being reviewed and/or the panel, or too expensive or extended;
- ii. At the other extreme, it might become too light to be credible – e.g. too cursory or superficial a treatment, or having insufficient engagement of the panel members to warrant the claim of a properly informed peer assessment;
- iii. It might become, or be perceived as, too “cosy” an assessment among professional evaluation peers, lobbying for this professional function or exercising too much “professional courtesies” in its assessments; or
- iv. Because it requires access to sometimes-sensitive information and leads to important assessments of the quality of the work of organizations and teams, each review will encounter and have to manage different views on process and substance, and could become bogged down in disputes at various levels.
- v. It might be perceived as a donor-driven exercise to justify decreased voluntary contributions to the organization concerned if the conclusions of the review are negative;
- vi. At the other extreme it might be perceived as an exercise conducted to provide justification for higher voluntary contribution if the review is very positive.

The first three sets of risks identified above have each been encountered and tested in the pilot reviews, and the framework outlined in this paper has been designed to include features to mitigate each of them to the extent possible. At the same time, it will be important for all parties embarking on such a review to register and discuss these risks explicitly at an early stage, and build any necessary risk-mitigation measure into the specific review plan.

The fourth risk – that of serious disputes which might come to threaten the integrity or completion of the review – also needs to be anticipated and specifically managed. It should be stressed that differences of approach and interpretation are intrinsic to any such exercise, and that the open “dialogue and interactive investigation” in the process are a major part of its value. Nonetheless, even in very well institutionalized peer review systems, intractable conflicts can arise among participants at various levels. In such cases – always recognizing that the process is a non-coercive one – the normal recourse is to call on the body sponsoring the review system to help resolve or manage the conflict. However, since that body in this case does not have any legal status, it would be preferable to invite the governing body of the organization

concerned to initiate a process of mediation. This body may seek advice or support from the co-chairs of the Joint UNEG-DAC Task Force and mediation efforts could include appointing one or more arbitrators to advise the governing body on how to proceed. It is recommended that this contingency should be specifically recognized and provided for in the agreed peer review plan between the panel and the host organization.

The last two risks will need to be confronted in the mixture of panel members, where restraint should be applied in appointing members from donors that have a high financial stake in the organization or donors that are in the process of reconsidering the level of their voluntary contributions.

Focus of the peer reviews of evaluation.

The professional peer review would focus on the evaluation function within the organization, take its central evaluation unit as a starting point and working towards understanding how evaluation operates at the various levels, in order to review the quality of the function in light of the objectives of the organization and the appropriate international standards. Specific aims would be:

1. To assess structural aspects of how the evaluation function operates in the organization. Depending on the organization reviewed, this should look at the evaluation function in the perspective of how the organization can better report on and increase its effectiveness in reaching its goal.
2. The peer review should look at policy issues, starting with the evaluation policy of the organization and other related policies, as well as issues of planning, budgeting and coverage.
3. It should address various relevant levels in the organization, such as the governance level (board, council), central management functions and the central evaluation unit, decentralized evaluations both in headquarters and in field offices, and outsourced evaluations in interventions.

It should look at existing capacity and need for capacity strengthening, use of relevant and appropriate evaluation methodologies, and the quality of the evaluations undertaken, in preparation, implementation and reporting. Another key focus should generally be to look at collaboration in evaluation with local partners and stakeholders, as well as harmonization and coordination with other external partners.

The framework for assessment

A core assessment question.

The professional peer reviews of evaluation are to be centred on a “core assessment question”:

‘Are the agency’s evaluation function and its products: independent; credible; and useful for learning and accountability purposes, as assessed by a panel of professional evaluation peers against international standards and the evidence base.’

As reflected in this question, the approach and methodology hinges on using recognized normative frameworks and *clustering* the treatment of the many issues *under three crucial criteria for evaluation*, specified as:

- J. *Independence* of evaluations and evaluation systems. The evaluation process should be impartial and independent in its function from the process concerned with the policy making, the delivery, and the management of assistance. A requisite measure of independence of the evaluation function is a recognised pre-condition for credibility, validity and usefulness. At the same time, each review should bear in mind in that the appropriate guarantees of the necessary independence in a particular organization will differ according to the nature of its work, its governance and decision-making arrangements, and other factors. Moreover, most organizations aim to encourage the active application and use of evaluations at all levels of management, meaning that systemic measures for ensuring the necessary objectivity and impartiality of this work should receive due attention.
- K. *Credibility* of evaluations. The credibility of evaluation depends on the expertise and independence of the evaluators and the degree of transparency of the evaluation process. Credibility requires that evaluations should report successes as well as failures. Recipient countries should, as a rule, fully participate in evaluation in order to promote credibility and commitment. Whether and how the organization’s approach to evaluation fosters partnership and helps builds ownership and capacity in developing countries merits attention as a major theme.
- L. *Utility* of evaluations. To have an impact on decision-making, evaluation findings must be perceived as relevant and useful and be presented in a clear and concise way. They should fully reflect the different interests and needs of the many parties involved in development co-operation. Importantly, each review should bear in mind that ensuring the utility of evaluations is only partly under the control of evaluators. It is also critically a function of the interest of managers, and member countries through their participation on governing bodies, in commissioning, receiving and using evaluations.

Normative framework

In order to test the organization's evaluation function against these three criteria in the necessary depth, accepted normative frameworks are required as instruments for applying the assessment. Depending on their applicability to the organization being reviewed, one of several sets of accepted and consistent standards should be applied to provide such a normative framework:

- For UN organizations, the Norms and Standards produced in 2005 by the UN Evaluation Group (UNEG);
- For the participating Multilateral Development Banks, the good practice standards and template for independence of their Evaluation Cooperation Group; and
- For bilateral organizations, standards found in the OECD/DAC Principles for Evaluation.

The assessment against these standards can be supplemented by other questions judged relevant by the peer Panel (which should then be clearly specified in its report).

In order to proceed with applying the relevant norms and standards and ensuring proper coverage of the three main aspects, the norms and standards need to be broadly re-grouped and checked under those three headings. In the pilot studies, this process was carried out for the UNEG Norms and Standards and, in spite of some overlaps and repetitions¹⁴ – inevitable with any such type of categorization – they did ultimately provide an adequate basis for organizing each review, its report, and the communication of its results. Similar work was done for the DAC Principles. Even taking account of the difficulties, the clustering of these norms under sharp and understandable headings such as the “independence, credibility and use” of evaluations clearly also helps in providing some thematic shape and comparability at an appropriate level. To avoid future professional peer reviews of UN organizations having to repeat this sorting and categorization work needlessly, such review teams are referred to the documentation of the UNDP and

¹⁴ The UNICEF pilot peer review report was explicitly critical of the difficulties caused by the lack of a neat “fit” between the UNEG Norms and Standards and the categories of independence, credibility and usefulness, and recommended the search for a neater assessment approach, as well as some streamlining of the Norms and Standards. However, the Joint Task Force has not yet found any alternative assessment framework that would capture both the accepted standing of the Norms and Standards, and the useful clustering under these three important headings. It has also concluded from the two pilots that the difficulties are not insuperable and will ease somewhat in subsequent reviews. The goal of making suggestions to streamline the various sets of norms and standards in the field is a shared one, although it will require some more testing of their use and, of course, not just for peer reviews.

UNICEF peer reviews to be able to extract readily the relevant materials.¹⁵

The experience of the pilot reviews suggests that it may normally be appropriate to apply a combination of indirect (evaluation system) and direct (sample evaluation product) assessments at different levels. Although there are not yet single, internationally accepted standards for assessing the quality of evaluation products, a number of useful models are already being tested and applied at different levels. For example, a good number of the UNEG standards bear directly on the quality of evaluation reports, and in March 2006, the DAC Evaluation Network released a set of DAC Evaluation Quality Standards (on processes and products) for a 3-year test phase application. A number of organizations have also been developing and applying their own systems in these areas. Peer review teams and their host organizations should thus consider and decide at an early stage on the quality standards that should be applied, in an experimental spirit, for the review in question, and then document and report on the experience with their use.

In the same experimental vein, it should also be recognised that these peer reviews may be one of the most important continuing general tests of the usefulness of the prevailing norms and standards for evaluation, and they should thus be prepared to suggest ways to strengthen the norms on the basis of these experiences.

Given the central importance of the Normative Framework to the whole professional peer review process and its results, experience confirms the importance of the specific framework being agreed between the peer review team and organization at an early stage. As the Lessons from the UNDP pilot review stressed, "...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."

Roles and responsibilities in the reviews

The Peer Panel

The expertise and commitment of the peer panel itself, together with the cooperation of the organization being reviewed, is essential to the successful conduct of the peer review. Given the likely demands of other responsibilities on most panel members, the demands of this work need to be well-organized and prepared, and highly experienced and skilled consultants are also likely to be required.

¹⁵ In summary, for UN organizations, indicators of independence are broadly covered by UNEG Norms N6.1–N6.5 and amplified in the relevant Standards. Indicators of credibility are mainly treated in UNEG Norms N5.1–N5.3, N8.1, N9.1–N9.3 and N11.1–N11.5 and amplified in the relevant Standards. Indicators of utility are mainly treated in UNEG Norms N2.6, N1.3, N8.2, N10.1, N10.2 and N12.1–N12.3 and amplified in the relevant Standards

In selecting participants for the panel, which should be restricted to five or six participants, experience confirms that the major criteria for selection of the group should include a combination of the following attributes:

- i. Independence from the particular organization being assessed;
- ii. Professional evaluation expertise;
- iii. Understanding of the context and use of evaluation in development cooperation and multilateral organizations;
- iv. An acceptable gender mix;
- v. Participation from multilateral and bilateral agency/ies;
- vi. Participation from country/ies receiving assistance, including those with evaluation responsibilities;
- vii. Participation from independent evaluation experts; other research fields, oversight disciplines or knowledge sharing expertise; and
- viii. Capacity to deal with senior management and governing body levels.

It is recommended that either one bilateral evaluation office represented on the peer panel or several members of the peer panel take the *lead responsibility* for:

- Managing the process and either directly providing logistical and secretariat services, or contracting in relevant support;
- Being willing and ready, through representation on the multilateral organization's board, to engage with the agency's senior management and governing board to actively encourage their interest, and the consideration of recommendations and possible action following the assessment process; and
- Ensuring that the key conclusions and possible lessons from the review are communicated to the wider UN Evaluation Group and DAC Evaluation Network membership and that the reports and supporting evidence are easily accessible to members and others.

Although the peer panel takes overall responsibility for the process, together with the organization reviewed, the most intensive direct roles of panel members would normally be concentrated near the beginning and end of the overall process:

- First, in clearly defining the framework and *modus operandi* to be proposed for the particular review, with the organization being reviewed¹⁶ (phases 1–3 below); and

¹⁶ To ensure mutual clarity and smooth working relationships, the panel and its consultants should normally take responsibility for briefly setting out the agreed plan for the specific peer review on paper for reference as required. It should reflect the application of this framework and specific adaptations.

- Later in validating the evidence and findings, bringing together the judgements, conclusions and recommendations of the review, presenting and following up the results. (Phases 5 & 6 below).

The major part of the intervening data-gathering and analysis work (phase 4 below) would normally be expected to be delegated to carefully-selected consultants,¹⁷ responsible to the panel and working within the agreed framework.

The evaluation unit of the organization under review

The evaluation unit is likely to be the internal initiator, principal organizer, and substantive collaborator on the review within the organization concerned, although to the extent that a governing body or senior management are willing and able to take more active roles in the first two areas this is likely to have major benefits to the credibility and usefulness of the review. The evaluation unit (or other principal partner to the review from the organization) will be undertaking the responsibility for considerable work in the following areas:

- Informing and obtaining the necessary endorsements by senior management, and if possible the governing body, to conduct the review and assure the necessary cooperation;
- Up-dating or preparing an initial self-assessment of the evaluation function in the organization;
- Considering and agreeing to a specific design and plan for the review;
- Facilitating all necessary access to its own data and internal interviewees, and assisting in identifying and securing similar access throughout the rest of the organization;
- Helping ensure the presentation of the results to senior management and the governing body, and its dissemination throughout the organization; and
- Contributing its views on the experience to the Joint Task Force, and providing follow-up reports (after one and two years) on actions taken and other impacts of the review.

¹⁷ The 2005 approach and methodology for the two pilot studies had also envisaged the options of panel members or their colleagues carrying out this work or, if agreed, the host organization itself. In the event, two-member teams of consultants from different countries were retained to carry out this work. In addition to a high level of skills in evaluation, synthesis and communication, the selection of consultants should also reflect key requisites for the panel, such as independence from the particular organization being assessed, a very good understanding of the context and use of evaluation in development cooperation and multilateral organizations, and the ability to work well in this capacity with the host organization at various levels as well as the panel. Including consultants from country/ies receiving development cooperation is an important objective. The two pilot studies had one two-male team and one two-female team – whether or not a gender mix is always possible, good knowledge of gender issues should be a requirement on the consultant team. It is important that the consultant team be in place early in the process to acquire an excellent grasp of the panel's approach and the arrangements with the host organization.

Phase-by-phase organization of a professional peer review process (illustrated with typical steps and sequences).

Phase 1: Initiating and organizing the review.

- i. Invitation by the organization (preferably the governing body or senior management) to conduct the professional peer review. It should be noted that such peer reviews are likely to be best planned and scheduled to mesh with important decisions, reviews and/or relevant events;
- ii. Volunteering (ideally in some rotational order) by a representative of a lead peer organization, willing to offer direction, coordination and financial support, backed by a consortium of other peer organizations also willing to participate. (In addition to the substantial in-kind contributions required of them, some organizations being reviewed might in some cases want to take some share in the direct costs.);
- iii. Endorsement by senior management of the host organization and if possible by its governing body. (At the very least, the latter should be informed and engaged at the outset and at key milestones in the process. Demonstrating benefits from other experiences may be helpful in stimulating interest and demand by both senior management and governing bodies);
- iv. Constitution of the peer panel by the lead organization and consortium members, in consultation with the host organization concerned and the Joint DAC/UNEG Task Force;
- v. Selection and appointment of consultant/s to support the peer panel;
- vi. Preparation and circulation of some form of self-assessment of its evaluation function by the organization concerned (e.g., for UN organizations, up-dating of the responses to the 2005 UNEG self-assessment checklist, or an appropriate equivalent for other organizations);
- vii. Thorough familiarization, normally by the review consultants, with the evaluation function in the context of the organization's mission, governance, policies, structures, budget, staffing and operations. It is important to first try to understand the organization, and how the evaluation function can best serve it, prior to looking in depth at the machinery.

Phase 2: Agreeing on the review methodology, work-program and mutual responsibilities with the host organization.

- i. The panel would normally receive intensive orientation briefing by its consultants on the organization and its work, and the results of its self-assessment of the evaluation function. This could be followed by direct meetings with the evaluation unit and possibly others to cross-check and verify this starting base;
- ii. Following the previous step, the peer panel should pursue a discussion and clear agreement with the evaluation unit on the core assessment question, the basic normative framework and methodology, the work-program and mutual responsibilities for the conduct of the review, including an outline of the succeeding phases, specifying anticipated data requirements, scheduling and arrangements for access to sources, interviewees, possible field missions, etc.
- iii. The panel provides guidance to consultants on its priority interests for the collection and analysis of data, and preparation of findings. Depending on the character of the organization's evaluation functions, data-gathering and findings might be guided by emphasis on centralized or de-centralized evaluation processes, appropriate and feasible samples of individual evaluations and their products, etc.¹⁸ An "issues paper" could be the appropriate form to gather the evidence and analysis in this stage of the process.

Phase 3: Collecting and analysing the data, shaping findings against the normative framework adopted

- i. Normally carried out in the main by the panel's consultants, in consultation with the evaluation unit and others, this phase would be likely to include: extensive review of the relevant files and documentary information base to be gathered and analysed in line with the core assessment question and the normative framework; key informant interviews, focus groups and field missions around systems and processes and/or particular evaluations, which are used as reference cases;
- ii. Preparation of the base of evidence and findings for the panel's review.

¹⁸ The two pilot reviews illustrated the need to adapt the approach. The review panel for UNDP focused on its centralized evaluation function and office. To make the informational foundations as concrete, focused and systemic as possible within the scope of the assessment, it selected a small group of recent evaluation processes and products of different types carried out by the Evaluation Office as reference cases, taking into account a combination of several criteria. In the case of UNICEF, with a highly de-centralized evaluation system, the methodology placed considerable emphasis on an intensive examination of evaluation within a particular country program, Ghana in active engagement with Ghanaian and other partners in that country.

Phase 4: Validating the evidence and findings, shaping conclusions and recommendations.

- i. 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 from the evaluation unit, the Panel will conduct wider interviews with a number of Senior Managers and Executive Board members concerned. The reference cases will serve as one important focus;
- ii. Integrating the results and insights from these interviews, the Panel will complete the triangulation, refinement and confirmation of its base of evidence and findings. On that basis, it will then move into the “judgement phase” following the following steps: panel agrees on its main frameworks for judgements in relation to the Norms and main aspects of the assessment; panel considers and debates the evidence and findings, and arrives at its draft conclusions and recommendations, agrees on draft report;

Phase 5: Presenting and discussing the review report, dissemination

- i. Draft report serves as the basis for the Peer Assessment meeting with the host evaluation unit to consider the results in depth;
- ii. Panel carries out final review and any necessary revision to the draft report;
- iii. Panel transmits its final report to the host evaluation unit, providing a final opportunity for review and reflecting dissenting views or other key responses;
- iv. Panel presents the Peer Review Report to the organization, ideally at the Governing Body and accompanied by a Senior Management Response.

Phase 6: Following up

Panel and organization reviewed report, ideally together, to the Joint UNEG DAC Task Force on lessons learned from the exercise, and its results over time.

Time and resources required

There will always be too many variables to allow any firm across-the-board estimates in advance for the time and resources needed for professional peer reviews of evaluation in different organizations (e.g. size and complexity of the organization of the organization and its programs; locations of panel members; calendars of work for the host organization and panel members, amounts of field work required, etc.)

At the same time, it may be helpful – on the basis of experience to date and discounting the up-front “development costs” of the first two “pilots” – to provide at least a *general sense of the minimum time and resources*

likely to be needed to carry out a peer review along the lines outlined in this document. The general estimate of elapsed time required would be expected to be at least six months from inception to reporting, but should not be more than nine months. Over such a period, panel members would each be required to devote at least 10–15 days of work, and the lead organization/s at least an additional 10–15 days of supporting and logistical work. The panel's strong supporting advisors or consultants would probably require a total of at least 60 days work. In the host organization, meanwhile, the evaluation unit will probably be required to devote at least 30–35 days of professional time, and significant administrative facilitation, while their colleagues elsewhere in the organization will be called upon for the time for interviews, assisting with field missions, etc.

Planned monitoring and assessment of experience

The Joint Task Force DAC/UNEG will continue to work on the following:

1. conducting a quick joint review of experiences following each peer review – as standard procedure;
2. stocktaking once every two years on the basis of such reviews;
3. updating of the approach and methodology (described in the current framework) on the basis of this periodic stocktaking.

Key references

1. Peer reviews of evaluation in international organizations, A joint initiative of the DAC Evaluation Network and the UN Evaluation Group, Draft August 9, 2006
2. Peer Review of Evaluation Function at UNICEF, Peer Panel Observations and Recommendations, June 22, 2006
3. Peer Review of UNDP Evaluation Office, Key Issues and Lessons Identified, 17 March, 2006
4. DAC Evaluation Network, A New Approach to Assessing Multilateral Organisations' Evaluation Performance, Final Draft. June, 2005
5. Peer Review: United Nations Development Programme Evaluation Office. December, 2005, DANIDA et al.
6. Peer Review of the Evaluation Function at United Nations Children's Fund (UNICEF). May, 2006
7. UNEG Norms and Standards
8. DAC Principles for Evaluation, and up-date
9. TORs for Drafting a Framework for Future Reviews

Appendix 3

Persons interviewed

(in person or by phone interview)

July and September missions in Rome

WFP staff

Amir Abdulla	Regional Director, Southern Africa Bureau
Stephen Anderson	Assistant to Director of Operations
Kojo Anyanful	Director, Internal Audit
John Aylieff	Emergency Coordinator & Director ODA
Bill Barclay	Head, Common Monitoring and Evaluation Approach
Luay Basil	Senior Regional Programme Adviser, ODC Bureau (Cairo)
Paul Buffard	Head of Programme, Sudan
Torben Due	Director, Programme Management Division
Jean-Jacques Graisse	Senior Deputy Executive Director
Wolfram Herfurth	Regional Programme Adviser, ODC Bureau (Cairo)
Deborah Hines	Chief, Performance & Reports, Policy, Strategy & Programme Support
Alan Jury	External Relations Division
Al Kehler	Senior Regional Programme Advisor, East and Central Africa Bureau (Uganda)
Adnan Khan	Director, Oversight Services
Paul Larsen	Chief of Staff, Office of the ED
Ramiro Lopez da Silva	Director, Transport and Procurement Division
Ute Meir	Deputy Chief, School Feeding Service
Christine Van Nieuwenhuysse	Deputy Regional Director, Western Africa
Daniela Owen	Deputy Director, Sudan
Kofi Owusu-Tieku	Senior RBM Officer
John Powell	Deputy Executive Director

Elvira Pruscini	Head of M&E, Sudan
Susana Rico	Regional Director, East and Central Africa Bureau
Stanlake Samkange	Director, Policy, Strategy & Programme Support
Ram Saravanamuttu	Programme Advisor, Operations
Suresh Sharma	Director Change Management Unit
Romain Sirois	Deputy Country Director, Haiti
Paul Turnbull	Chief, Technical Programme Support Services, Operations
Hildegard Tuttinghoff	Programme Adviser, Operations
Darlene Tymo	Senior Project Coordinator (SENAC), Emergency Needs Assessment
Alexander Voccia	Policy Adviser (Consultant), Policy, Strategy & Programme Support
Jamie Wickens	Associate Director Operations

WFP evaluation office (OEDE)

Caroline Heider	Director
Alain Cordeil	Senior Evaluation Officer
Michel Denis	Evaluation Officer
Aurelie Larmoyer	Evaluation Officer
Julian Lefevre	Chief Evaluation Officer
Anne-Claire Luzot	Evaluation Officer
Jeffrey Marzilli	Senior Evaluation Officer
Annemarie Waeschle	Senior Evaluation Officer

Others

Kees Tuinenberg	former Director, OEDE
Luciano Lavizzari	Director of Evaluation, IFAD
John Markie	Director of Evaluation, FAO
Rachel Bedouin	Senior Evaluation Officer, FAO
Sarah Longford	WFP Programme Adviser (on leave)
Francine Pickup	Humanitarian Evaluation Officer, OCHA
Mathew Varghese	Senior Evaluation Officer, UNICEF

Meta-evaluation related telephone interviews

Mohamed el-Kouhene	Country Director, WFP Yemen
Anneclaire Luzot	Evaluation Officer, OEDE
Sarah Gordon-Gibson	Deputy Country Director, WFP Niger
Manfred Metz	Berlin Team Leader, Yemen Evaluation (Independent)
Martin Steinmayer	Team Leader, Emergency School Feeding (Independent)
Aurelie Lamoyer	Evaluation Officer, OEDE
Jackie Kirk	Team Member (education) Emergency School Feeding (Independent)
Younes Bourfa	Team Leader Niger Evaluation
Jon Bennett	Team Leader Indonesia Evaluation
Hakan Tongul	Programme Officer and Team Member Myanmar

Hugh Goyder	Team Leader Tsunami RTE (Independent)
Julian Lefevre	Chief Evaluation Officer OEDE and Team Member Tsunami RTE
Alexander Voccia	Policy Advisor Georgia AAR
Annemarie Waeschle	Senior Evaluation Officer OEDE, Georgia AAR
Jeremy Shoham	Team Member Targeting (Independent)
Jeff Marzilli	Evaluation Officer OEDE, Targeting

Malawi

External interviewees (NGO partners and Government)

Paul Jones	Emmanuel International
Henry Khufi	Emmanuel International
Felix Mtonda	Oxfam
Henry Zaindi	Adra
Hyton Lefu	COOPI
Prisca Waluza	Malawian Red Cross Society
Christopher Singini	COOPI
Andrew Khumalo	World Vision International
Amos Zaindi	Goal
Debbie Ball	Goal
Joseph Zimba	CPAR
Jimmy Kadongola	CPAR
Olex Kamowa	CPAR
Lilian D. Ng'oma	Secretary & Commissioner Department of Poverty & Disaster Management Affairs
James Chiusiwa	Relief Coordinator, Department of Poverty & Disaster Management Affairs

WFP staff

Irene Kumwenda	Programme Assistant, School Feeding Unit
Ashwini Rai	Head of Logistics
RominaWaldemarion	Acting Head of Programme/VAM Programme Officer
Teresita Mosquera	Programme Officer M & E
Amos Misomali	Programme Officer (PRRO)
Ntolo Jailosi Ntolo	Data Manager
Duncan Kochelani	Programme Assistant Community Household Surveillance
Duncan Ndhlovu	Programme Assistant Food Security
Lazarus Gonani	Nutrition Programme Officer
Gertrude Kara	Programme Officer HIV/AIDS
Sarah Kaphamtengo	Assistant PO HIV/AIDS
Karla Hershey	Deputy Country Director
Domenico Scalpelli	Country Director

Johannesburg

WFP staff

Amir Abdulla	Regional Director
Thomas Yanga	Deputy Regional Director
Wanja Kaaria	Programme Officer M & E
Mutinta Hambayi	Programme Officer Nutrition & HIV/AIDS
Sonsoles Ruedas	Regional Programme Advisor

NGO partners or stakeholders

Neil Townsend	Regional Humanitarian Coordinator, OXFAM
Duncan Campbell	Regional Food Coordination Unit, World Vision
Francis Butall	Regional Food Coordination Unit, World Vision
Dan Maxwell	Deputy Regional Director (Livelihoods)
Karen Tibbo	Senior Food Security Adviser, CARE

Bangkok

WFP staff

Erika Joergensen	Deputy Regional Director
Parvathy Ramaswami	Programme Advisor M&E Focal Point
Rita Bhatia	Senior Programme Advisor Nutrition
Yvonne Forsén	Programme Advisor Emergency Assessment
Asaka Nyangara	Programme Advisor
Gerry Daly	Senior Programme Advisor Head of Programmes

Jakarta

WFP staff

Bradley Busetto	Country Director
Janne Suvanto	Head of Programmes
Carolyn Hardy	Procurement Officer
Yendri Adam	National Procurement Officer
Bassam Hana	Logistics Officer
Konendera Belliapa	Head of Logistics
Alan Brown	Private Sector Partnership Consultant
Kornelius Schiffer	Programme Officer Aceh
Thi Van Hoang	Programme Officer (regarding Georgia AAR)

Partners or stakeholders

Siti Halati	Director Field Operations and Data Management Helen Keller International
Adang Setiana	Deputy Minister Coordinating Ministry for People's Welfare
Pak Kresnawan	Ministry of Health

Vientiane

WFP staff

Christa Röder	Country Director
Sangsaath Vongkhamsao	Senior Programme Assistant

Partners or stakeholders

Adam Folkard	Consultant to Evaluation (Independent)
Leik Boonwaat	Country Director UNODC
Nigel Orr	Swiss Foundation for Mine Action
Joseph Huber	Swiss Foundation for Mine Action

Appendix 4

Background data for the Peer Panel members and advisors

Jock M. Baker is currently the Programme Quality & Accountability Coordinator for CARE's Emergency Group, which is based in Geneva, Switzerland. Since joining CARE in 2001, Mr. Baker has been one of CARE's leading technical voices for the agency's global relief and transition programs, supporting the development of policies and guidelines and building capacity through training and dissemination of good practice. He has led a number of thematic reviews of organizational policy in addition to participating in assessments and evaluations of humanitarian actions, among those a global study of CARE's work in conflict-affected settings, development and testing of environmental impact mitigation tools for use during disaster response, an evaluation utilisation study, and in addition led and managed a number of joint evaluations and After Action Reviews. Mr. Baker is a member of the American Evaluation Association, Steering Committees for various UN-led evaluations, and a founder member and former Chair of InterAction's Transition, Conflict & Peace Working Group. He represents CARE on various Humanitarian Accountability networks and, among other published works, is a contributing author/editor for the 2004 edition of the Sphere Handbook. Prior to joining CARE, Mr. Baker was an independent consultant following a career spanning over fifteen years as a UN staff member in successive assignments with WFP, UNHCR, UNDP and OCHA. Mr. Baker holds a BSc (Hons) in Biological Sciences from the University of Edinburgh and an MSc in Economics from the London School of Economics and Political Science.

Stefan Dahlgren is a social scientist by training and before joining Sida in 1987 he worked for around 20 years with research in housing policy and housing conditions in Sweden, most of the time at the Swedish National Housing Research Institute, and for some years as advisor at the Building Research Institute in Tanzania. He has been involved with evaluations around half of his time at Sida and with operations during the other half when posted in Vietnam and Afghanistan. From 1991 to 1994 he was head of Sida's evaluation department and has during the last four years worked mainly with evaluations of humanitarian issues, most recently being responsible for the LRRD part in the international Tsunami Evaluation Coalition (TEC) evaluation. In 2007 he became a member of ALNAP's Steering Committee.

Susanne Frueh represents the UN and the United Nations Evaluation Group (UNEG) on the Panel. She has over 21 years of work experience with six different parts of the UN and two International Finance Institutes and has been a member of the American Evaluation Association, a board member of the UN Evaluation Group and a steering committee member of ALNAP. Following positions in project and programme management in the field and headquarters with UNDP and UNOPS, she switched over to evaluations with the United Nations Capital Development Fund in 1995. She subsequently worked for WFP (OEDE) as senior evaluation officer from 1999–2002 and then served as head of evaluations for OCHA until May 2007. While at WFP, Ms Frueh worked on the M&E Guidelines, early results-based management initiatives, and the evaluation policies issued between 1999 and 2003. Ms. Frueh was one of the drafters of the UNEG Norms and Standards for Evaluation in the UN system. While at OCHA she was also involved in a number of inter-agency evaluations such as the Darfur RTE and the Mozambique RTE and developed the inter-agency RTE pilot jointly with UNICEF.

Ted Kliet is a senior evaluator of the Policy and Operations Evaluation Department, Ministry of Foreign Affairs, the Netherlands. His responsibilities are to carry out and report policy evaluations in the field of Dutch development cooperation and other fields of foreign policy. During the past 15 years he has managed several joint evaluations where the Policy and Operations Evaluation Department is leading and has participated in joint evaluations led by other development partners, among those the evaluation of Humanitarian and Reconstruction Assistance to Afghanistan 2001–2005 and the multi-agency evaluation of Humanitarian Support to Internally Displaced Persons. Also he chaired the management group of the Joint Evaluation of External Support to Basic Education (published in 2003). He was responsible for the evaluation of Dutch Humanitarian Assistance (published in 2006). He is currently

engaged with the Joint Evaluation of the Paris Declaration and the Joint Donor Evaluation of Managing Exit Strategies. Mr Kliest is a member of the OECD/DAC Network on Development Evaluation and is co-chair of the OECD/DAC-UNEG Task Force on Peer Reviews of Evaluation Functions of International Organisations. He has been involved with ALNAP since 1999 and was a member of ALNAP's Steering Committee. Mr Kliest was a member of the Peer Panel, which assessed the evaluation function of UNDP Peer Review report published in 2005.

Zenda Ofir has worked since 2000 as full-time evaluation specialist in diverse fields of development in 26 countries in Africa, Asia and Europe. She has experience at local, national, regional and global level and her nearly 40 clients to date include national governments and universities as well as bi- and multilateral organizations such as UNDP, UNIFEM, IDRC, IFAD, CGIAR and ILO. In 2005 she worked for the Executive of the World Conservation Union (IUCN) as Special Adviser on Knowledge Management. She recently spent five months as visiting professor at a national centre for the study of international cooperation in education located at the University of Hiroshima in Japan. Ms Ofir has served as President of the African Evaluation Association (AfrEA), as Vice-President of the International Organization for Cooperation in Evaluation (IOCE), as first ever Board member of the American Evaluation Association (AEA) based outside the USA, and as Co-Chairperson of the Third AfrEA Conference held in 2004 in Cape Town with 550 participants from 56 countries. She was a member of the International Expert Panel for the review of IFAD's evaluation methodology and is currently a subgroup coordinator for the development of guidance on impact evaluation for the three networks of bi- and multilateral donor evaluation offices (Network of Networks on Impact Evaluation – NONIE). In 2008 she will become a Member of the Standing Panel on Impact Assessment of the CGIAR Science Council. A South African with a PhD in Chemistry, Ms Ofir's early career included appointments as National Science Council grants manager and as Director of Research and International Affairs of the University of Pretoria.

Ian Christoplos, advisor to the Peer Panel, is a researcher at the Department of Urban and Rural Development at the Swedish University of Agricultural Sciences and an independent consultant. His work focuses on issues related to risk, humanitarian assistance, rural development and agricultural services. He has worked as a researcher and practitioner in both development cooperation and humanitarian assistance in Africa, Asia, Latin America and the Western Balkans.

Peta Sandison, advisor to the Peer Panel, has worked in the humanitarian sector since 1990 as a manager and consultant evaluator. She has a Master of Science in Land and Water Resource Management, a BA (Hons) degree in English and American literature and qualifications in management and coaching. She has worked as a senior field and headquarters manager for INGO humanitarian programmes in Africa and Asia for over 10 years and has extensive experience of working in complex emergency contexts. Since 1999 she has also worked as an independent consultant, carrying out evaluations, studies and internal reviews for UNICEF, UNHCR, the Red Cross Movement, DFID, the Disasters Emergency Committee and several international NGOs. She is the author of *The Utilisation of Evaluations* published in the ALNAP *Review of Humanitarian Action* published by ODI in 2006.

Peer Review EVALUATION FUNCTION_{AT} THE WORLD FOOD PROGRAMME_(WFP)

This is the third professional peer review of the evaluation function at a multilateral development or humanitarian organisation. It was carried out at WFP's request by a group of international evaluation experts. The review has assessed the strengths and weaknesses of the evaluation function at WFP along three criteria or dimensions: *independence*, *credibility* and *utility*. A number of measures for improvement are proposed in the report.

The professional peer reviews are joint initiatives by the OECD-DAC Network on Development Evaluation and the UN Evaluation Group (UNEG).

