Food and Agriculture
Organization of the United Nations

Evaluation Service
(PBEE)

AUTO-EVALUATION GUIDELINES

Version 1. 1

November 2003

“In pursuing my policy to further invigorate FAO as a centre of excellence, I place a particular emphasis on evaluation as a means of strengthening our capacity as a learning organization, especially for continuous improvements in the Organization’s programmes.”

Jacques Diouf, FAO Director-General
DG Bulletin No. 2001/33

“The most important purpose of evaluation is not to prove but to improve.”

Egon G. Guba

Table of Content

List of Acronyms

ADG	Assistant Director General
AE	Auto-evaluation
CP	Continuing programme activity
CSO	Civil Society Organization
DG	Director General
DSA	Daily Subsistence Allowance
DOs	Decentralized Offices
FIVIMS	Food Insecurity and Vulnerability Information and Mapping System
FOP	Forest Products Division
GILW	WAICENT / FAOSTAT Data Management Branch
GPA	Gender Plan of Action
MTP	Medium Term Plan
NGO	Non-governmental Organization
PAIA	Priority Area for Inter-disciplinary Action
PBE	Office of Programme, Budget and Evaluation
PBEE	Evaluation Service
PE	Programme Entity
PIR	Programme Implementation Report
PIRES	Programme Planning, Implementation Reporting and Evaluation Support System
PRA	Participatory Rural Appraisal
PWB	Programme of Work and Budget
SPFS	Special Programme for Food Security
SWOT	Strengths, Weaknesses, Opportunities and Threats
TORs	Terms of Reference
TP	Technical project
TS	Technical services agreement
WHO	World Health Organization

INTRODUCTION

The Director-General's Bulletin No.2001/33 on "Strengthening the FAO Evaluation System" and the Guiding Principles for Pre-Evaluation Monitoring, Annual Assessment and Periodic Auto-Evaluation of the Technical and Economic Programmes[1] introduced two complementary processes: annual assessment and auto-evaluation. Annual assessment is an annual exercise chiefly concerned with monitoring work plans and biennial outputs produced during the previous year. Auto-evaluation (AE) reviews programme achievements over a longer period – generally six years – and looks at a broader scope including results against planned outcomes and programme entity objective.

The aims of the new evaluation regime are:

a) To enhance programme managers' capacity to ensure the effectiveness, efficiency and relevance of programmes by facilitating (i) timely in-course corrective action and adjustment and (ii) assessment of programme achievements and results as a basis for deciding upon their future at critical points in the rolling MTP cycle;

b) To make assessment and auto-evaluation systematic and transparent, using a set of common criteria and procedures which can support programme planning and evaluation at the corporate level;

c) To contribute to the preparation of periodic accountability reporting, such as the biennial Programme Implementation Report (PIR) and other progress reports to management and the Governing Bodies; and

d) To provide a strong basis for independent evaluation by the Evaluation Service and external evaluators.

These aims cover both annual assessment and auto-evaluation, with some variation. Point c) for instance holds greater relevance for annual assessment than for auto-evaluation.

As far as evaluations are concerned and while each of them has its own particularities and areas of emphasis, they all essentially pursue the following three goals:

i. Programme improvement – Result-Based Management: analyzing the strengths, weaknesses and opportunities of a given programme leads to a set of recommendations on how to strengthen it and how much resources it should receive;

ii. Learning: evaluations may yield lessons that are valid over and beyond the evaluated programme, thus helping in the planning and management of future projects or programmes; and

iii. Accountability: public institutions and projects funded with public money must review and report openly about their contribution to the public welfare and to public goods within their mandate;

[1] Communicated by a memorandum dated 21 August 2002.

Typically, donors tend to place their emphasis on accountability, while programme staff and partners are more interested in programme improvement and learning.

Negotiating the Terms of Reference for an evaluation often involves trade-offs about where the main emphasis should be placed among the above three typical objectives. While underlining the classic role of evaluation as an accountability tool, the DG bulletin placed a strong emphasis on evaluation as a way to foster learning and continuous improvements in FAO programmes.

Programme managers may be familiar with the independent evaluations run by PBEE but less so with the auto-evaluation approach. Independent evaluations have advantages, such as being usually more objective than internal reviews. But they have drawbacks as well: they tend to be costly; the evaluators may not be totally familiar with the context and specificities of the work under review; their recommendations are not always accepted by the evaluated programme; and immediate lesson learning and internalization of conclusions are not assured. External and extractive evaluations, performed with accountability as their main goal, tend to be better at studying the past than at shaping the future of the evaluated programmes, due to a lack of ownership of the process and results.

In other words, independent evaluations are strong on the accountability side, but weaker on programme improvement and learning. In contrast, self (or auto-) evaluations tend to be strong on the improvement, forward-looking side but weaker on accountability. It is therefore important for FAO to combine the two approaches. Independent evaluations run by PBEE will continue to provide accountability and transparency to management and donors, while AE is intended to facilitate programme improvement and lesson learning.

In addition, it is essential that AE processes and reports be credible and transparent, and so perceived by our Governing Bodies and within FAO itself. In order to secure some objectivity, AE must be supported by external inputs and feedback, such as the review of evaluation reports by panels of external experts, the use of external consultants in conducting the evaluation, and the survey of our partners' and clients' views on the services we provide.

Auto-evaluation will focus initially on the technical and economic programmes (Chapter 2 and Major Programme 3.1). Non-technical programmes will start to be addressed during the next biennium (2004-2005). In particular, all technical projects (TPs) and continuing programme activities (CPs), as well as all PAIAs, will be subject to auto-evaluation during the six-year period 2003-2008. Some forty TP and CP programme entities should be evaluated during the initial two years (2003-04), together with a couple of PAIAs. While the selection of specific individual entities will be made by the Departments and Divisions in consultation with PBEE during the last quarter of 2003, the guiding rule is that all TPs coming to end during the biennium 2002-03 must be covered in the first group of Auto-evaluations.

Auto-evaluations may vary in the scope of coverage of individual programme entities. The minimum scope would be the auto-evaluation of one single programme entity, the maximum being that of a few related programme entities, or conversely, of an entire Programme. Responsibilities for oversight and management will depend on the scope of each specific auto-evaluation.

All AE reports, once cleared by the relevant Division Director(s), should be submitted to the ADG of the Department for approval. Regional Representatives should also be copied AE reports concerning those entities to which their Offices have contributed, so that they

may provide their own comments to the relevant ADG. Following the ADG's review of the report and of comments by Regional Offices, a final version will be produced and forwarded to all parties involved including Regional Representatives and PBEE, together with a note recording the ADG's decisions about the future of the programme entity(ies).

To be successful, the introduction of this innovative evaluation approach will require a new set of skills from programme managers. It is hoped that the present guidelines, coupled with an ongoing training programme, will help familiarize Regular Programme staff with the auto-evaluation modality and build up evaluation skills throughout the Organization.

PART I: CONCEPTUAL FRAMEWORK

The starting point in any evaluation is usually the evaluated project as described in the project document, i.e. as planned, and a good part of what an evaluation does is to compare actual achievements with objectives and targets. An overview of the conceptual framework underlining the New Programme Model and of its terminology are provided here to help ensure that the AEs performed in the Organization undertake this comparison in a similar fashion, and that when "auto-evaluators" across the house are referring to such things as output, outcome and impact, they actually talk of the same things.

A. Definitions

Definitions for each of the New Programme Model levels are summarized below:

- Input: financial, material and staff resources used or allocated to undertake an activity.
- Activity (or tasks): any portion of work by FAO staff or consultant, as opposed to the result of such work, which constitutes an output.
- Biennial Output: product or service, or group of similar products and services, produced by FAO during a biennium. Most often constitutes an element in a larger deliverable, for instance a publication in a series, a module in a database, or a particular training course in a larger capacity building effort.
- Major Output: significant product or service, or more often a group of related products and services, produced by FAO over one or several biennia and delivered to specific users. Normally composed of several, interrelated biennial outputs.
- Outcome: the way major outputs or biennial outputs are used, and the immediate result of this use. A short step towards longer-term impact. Not something FAO itself produces, but rather the way FAO outputs are "taken on" by its most immediate audience and clients.
- Programme Entity Objective: a statement of the benefits expected from the programme entity, beyond the immediate use to which the major outputs are put (i.e. beyond intended outcomes). A step in the line of causality between intended outcomes and the rationale. It can be conceptualized as a "second-degree" outcome. As for outcomes, it is not something FAO itself produces.
- Programme Entity Rationale: a statement of the reasons for the programme entity, including the development problem to be addressed; the ultimate beneficiaries; why it is reasonable to believe that achievement of the programme entity objective will deliver a benefit to or for ultimate beneficiaries; and why is there a priority for FAO in carrying out this work.

The PE rationale therefore contains a summary description of the intended impact of the programme entity. Impact is usually defined as the long-term consequences of any endeavor, positive or negative, intended or not. It takes years to develop, and is most often quite complex in nature. This sort of long-term impact is *not* intended to be verified by auto-evaluations. Impact assessment usually calls for quite elaborate evaluation techniques, better left to professional evaluators in the context of external evaluations. The scope of investigation of AEs should normally stop at the level of achievements

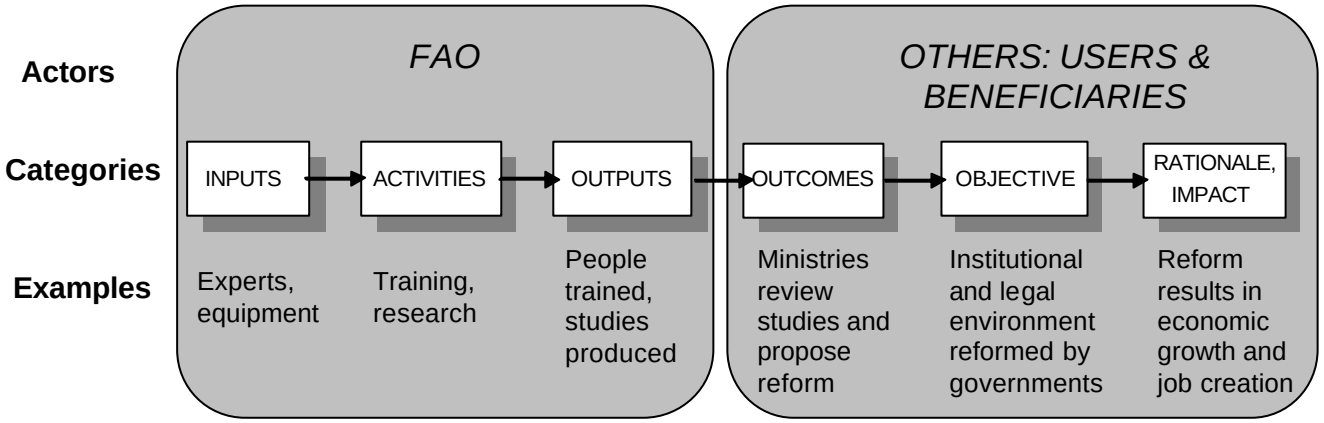
against the programme entity objective, complemented by a brief analysis of whether it is reasonable to expect that the programme entity will have a positive impact.

The rule of thumb to differentiate an output from an outcome is that your output is the result of your activities, something that will remain and be used once your work is completed, while your outcome is what somebody else, usually outside FAO, does with your output, how he/she combines it with other tools or services to produce his/her own output, for somebody else to use further down the line (see Figure 1).

B. Cause-to-Effect Relationships in the Normative Programme

The New Programme Model, as in any planning method, postulates some level of predictability and causality. There are such things as (admittedly complex) cause-to-effect relationships in the biological, economic and social fields. The hierarchy of results in Figure 1 is in fact a sequence of cause-to-effect relationships, each symbolized by an arrow, from activities to outputs, from outputs to outcomes and so on. Unfortunately, none of these cause-to-effect relationships is guaranteed to happen: some activities are unsuccessful, some outputs never used. Which is precisely why it is important to verify progress along the entire sequence through monitoring and evaluation.

Figure 1: Illustrative Sequence of Development Results



Social or economic cause-to-effect relationships are more complex than physical ones. Similar causes may have different effects, either by chance or because of scores of uncontrolled factors ("externalities"). The relationship is often of a supplier-client or teacher-learner nature: exchanges of knowledge, services and goods among free and largely unpredictable agents – hardly the reliable tick-tock of a mechanical clock. We should therefore hold no expectation of hard-and-fast attribution. The most one can claim in the development field and particularly in knowledge-transfer, is to have *contributed* to a development result, together with other actors and factors.

FAO normative activities generate information products and provide a forum for discussing and developing new approaches. They produce guidelines, training courses, methodologies, statistics, databases, software, publications, CD-Roms and the like – i.e. information products – and facilitate the circulation of information through networks, conferences and consultative bodies.

They do so in order to *improve decision making* at various levels: international (e.g. support to international conventions and initiatives), national (e.g. support to policy development), or local (e.g. support to field projects). The rationale is that a well-informed decision making process results in value-added services to the general public and/or improved management of public goods such as environmental resources, world trade or genetic resources. The relation between normative activities and ultimate beneficiaries is therefore indirect: it goes *via* improved decision making by other development actors, such as governments, donors, NGOs, development projects and the private sector.

Furthermore, the normative programme does not normally reach decision makers directly. Decision makers in governments and donor agencies are often overworked. They are not likely to surf the internet and assemble interesting data for a detailed study. They do not attend many training courses, nor do they read much technical literature. Decision makers in development projects are under no-less tight schedules, and are so numerous and dispersed that it may not be cost-effective to try and reach them directly.[2]

To influence decision makers on a global scale, the normative programme needs to channel its information through a well-targeted, closer and more approachable audience composed of private sector analysts, technical and policy advisors in line ministries and in donor institutions, delegates in international fora, academics and teachers, trainers and training institutes, civil society lobbyists, demonstration projects, and last but not least, the national, international and specialized media, all sorts of people and institutions whose role is to locate valuable information, summarize it, and showcase it to a larger audience including decision makers.

In the guidelines and templates for the MTP 2004-09, those who directly access outputs from the FAO Regular Programme and convey it to decision makers were called *primary users*, while the decision makers at the international, national or local levels were called *secondary users*. The people ultimately benefiting from improved decision making were called *end beneficiaries*.

This terminology was perhaps counter-intuitive, in that the adjective "primary", used in the sense of "the first in a sequence of events", also conveys a sense of the *most important*, while "secondary", meaning the "second in a sequence", also connotes a *less important* group. Evidently "secondary users", i.e. high-level decision makers or front-line development projects, are in fact more important (decisive) than "primary users", i.e. opinion makers, advisors, trainers and so forth.

This conceptual framework is summarized in Figure 2. It is clearly a simplification of all sorts of complex modes of play. In real life there may be less or more than three levels of users; users may interact in more complex manners (e.g. feedback loops) than in the model. All models are simplifications and it is important to keep them practical.

[2] It is possible for a programme entity to provide direct support to a few field projects, but such an approach can only be effective in the long term if these are pilot projects, demonstrating a given method so that it is in due time replicated by other programmes. In this case, the pilot projects supported directly by the normative programme have a dissemination function. One obvious example of such an approach is the SPFS.

Figure 2: MTP Terminology and Typical Modus Operandi

Goal Structure	Stakeholders Terminology	Examples of Stakeholders
Rationale	End beneficiaries	Farmers, fisherfolk, private sector, general public, the environment (benefits present and future generations)
		↑ *Legal framework, services* ↓
PE Objective	Secondary Users	Decision makers at the national level, policy-makers, donors, field projects, service providers, extension workers
		↑ *Knowledge, recommendations, motivation* ↓
Outcome	Primary Users	Opinion makers, technical advisors in governments & institutions, researchers and analysts, pilot projects, the media, international consultative & legislative bodies, trainers
		↑ *Data, information, analyses, technical fora, advice* ↓
Major Output	FAO and partners	Regular Programme staff, consultants and colleagues in partner institutions (e.g. WHO for CODEX)
Biennial Output		

C. Consequences for Auto-Evaluation

Auto-evaluation was conceived as a mandatory review of implementation progress towards the results planned in the MTP. An integral part of the Result-Based Management approach encapsulated in the New Programme Model, auto-evaluation is primarily concerned with the documentation, description and assessment of produced results as compared to the MTP. Three caveats are in order here.

Firstly, the MTP often uses resource scenarios that exceed the actual resource levels made available to the Organization. Any comparison between programme achievements and MTP targets should take this discrepancy into account.[3]

[3] The PIRES MTP and PWB applications will provide a good picture of the progression of usually declining resource allocations, from the MTP stage to the PWB. Actual allocations may be obtained from the concerned Divisions.

Secondly, in an evaluation as in daily life, what lies far away from us is more difficult to perceive that what is nearer. The difficulty to collect data rises as one moves along the user chain and further away from FAO. Our outputs, being under our direct control, are fully known to us. If we want to better document their production, all we need to do is ask a few FAO staff and partner organizations, a group of people that is easy to locate and on whose collaboration we can usually count.

Documenting outcomes will require that we survey or interview our primary users, who may be more difficult to locate than FAO staff and less keen on setting time aside to respond to our questions. Secondary users are even more difficult to access than primary users. As for surveying end beneficiaries at the grass-roots level, this would appear impossible, at least if one wants to relate their perceptions to the FAO Regular Programme. This is one of the reason why it would not be realistic to assess long-term impacts – i.e. achievements at the Rationale level – in the framework of auto-evaluation.

Thirdly, a static description of *what* results are should be complemented with a more dynamic review of the *process*, of *how* these results were generated in order to propose forward-looking recommendations. Knowing what our outputs are and what use they were put to (outcomes) will not necessarily tell us much about how the real or potential users came to know of our products or services, what it was that they liked or disliked in our outputs, and why they decided to use them or not. Only when we can answer these sorts of questions will we be in a position to recommend specific changes in the design or dissemination of our products and services.

Similarly, the outputs of the FAO Regular Programme are influenced by a number of factors bearing on implementation, be they positive or negative. The most important implementation constraints should be identified in the framework of auto-evaluation so that recommendations can be put forward to try and solve them.

In summary, an auto-evaluation should primarily be concerned with:

- a qualitative and quantitative description of achievements at the output, outcome, and objective levels; and
- an analysis of those processes leading to the production of these results, including implementation constraints and opportunities, and how outputs were disseminated.

Another important consequence of the conceptual framework presented in the preceding pages is that, since the sequence of development results displayed in Figure 1 occurs *over time*, the sort of results that one can expect from, and therefore evaluate in, a given initiative depends to a very large extent on the *maturity* of the said initiative. As a crude rule of thumb, outputs can start to be produced in a couple of years after project inception, outcomes need at least three to five years to develop, and any effect at the PE objective level will normally take place five to seven years after project inception. Consequently, a mid-term evaluation of a Technical Project will tend to focus on outputs and a bit on outcomes, while a final evaluation should be able to assess a wider range of outcomes and achievements at the PE objective level.

PART II: PROCEDURES

A. What to Evaluate: Scope

All technical projects (TPs) and continuing programme activities (CPs) should be auto-evaluated during the six-year period 2003-2008.

Evaluating progress against a plan or a set of objectives is a natural and ordinary part of any endeavor and in this sense, auto-evaluation is not a new process. It is "business as usual" for many programme managers and will continue to be so while being given more emphasis and coherence across the Organization. It is therefore mandatory that auto-evaluations be planned by all organizational units as part of their Regular Programme of work, and be funded from the Regular Programme.[4]

Clustering several connected programme entities into one single auto-evaluation may increase the cost-effectiveness and coherence of the evaluation process, in particular when several programme entities address similar audiences or deal with similar issues. It may also have a cost in making the auto-evaluation process less directly relevant to each individual PE. As explained below, this is a decision for the concerned ADG to make. Suffice to say here that auto-evaluations will come in different shapes and sizes. The auto-evaluation of one single programme entity represents the minimum scope. It is not permitted to auto-evaluate one single major output. The maximum scope is that of a few related programme entities, or conversely, of an entire Programme. Departments should however be aware that the larger the evaluation, the more difficult it is to plan for and coordinate.

B. When to Plan and Perform an Auto-Evaluation?

During PWB preparation, each Department should prepare a schedule of auto-evaluations to be undertaken during the following biennium. The plan should be finalized in discussion with PBEE which has the organizational responsibility for ensuring that all programme entities are subject to systematic auto-evaluation over a six-year period. The final plan, once approved by the concerned ADG, should be forwarded to PBEE for information. This process will also help to ensure synergy with the programme of external evaluations.

For technical projects (TPs), auto-evaluation will tend to take place near or during the last year of their planned duration, which can never exceed six years. For continuing programme activities (CPs), an auto-evaluation will take place at least once in each six year period, at an appropriate time.

Cluster auto-evaluations may call for some flexibility as to how close to the end of their cycle TPs are evaluated.

For Auto-evaluations to be carried out during 2003-04, it is recommend that all TPs scheduled for completion during 2002-2003 should be covered, as well as about one-half of all CPs. The latter is suggested to avoid the crowding of auto-evaluations towards the end of the MTP period (2007), when most TPs will be completed, and hence will need to be evaluated at about that time.

[4] Financial support will be made available by PBEE during 2003 and 2004 only, to support selected auto-evaluations and facilitate the introduction of the auto-evaluation process in the Organization.

One important consideration to take into account when planning for AE is that auto-evaluation should feed into the preparation of the rolling MTP. Thus, PEs that are scheduled for substantive reformulation are good candidates for auto-evaluation some time before the related MTP preparation process.

C. Process and Roles – Who Does Auto-Evaluation?

1. Oversight and Management Roles in Technical Departments

Auto-evaluation processes must be truly participatory, involving at key junctures all the concerned technical units at Headquarters and in those Regional Offices that have contributed substantially to the work in question.

While Departments and Divisions will make their own arrangements concerning AE, some suggestions to this effect are presented below and summarized in Figure 3. Responsibilities for oversight and management should depend on the scope of each specific auto-evaluation. For auto-evaluations of a cluster of programme entities, ADGs would normally retain oversight responsibility and could nominate, among the concerned Service Chiefs, an overall AE manager for the coordination of the exercise. In contrast, the auto-evaluation of a single programme entity could be overseen by the concerned Division Director and managed by a Service Chief or PE manager.

All AE reports, once cleared by the relevant Division Director, should be submitted to the ADG of the Department for approval. PBEE should be copied the draft AE reports, as well as Regional Representatives for those entities to which their Offices have contributed, so that they may provide their own comments. The ADG will then review the report, taking into account comments received from PBEE and Regional Offices to decide how the report should be finalized.

Following the ADG's review, a final version will be produced by the AE manager and forwarded to all concerned, including Regional Offices and PBEE, together with a note recording the ADG's decisions about the future of the programme entity(ies) and including his response to the evaluation recommendations. The ADG will in particular decide on the future of the evaluated entities (extension, termination, modifications, formulation of new programme entities, etc.), help remove the main constraints identified during the evaluations, and re-allocate resources if required.

2. Involvement of External Inputs

An auto-evaluation should always involve some external expertise and inputs. This could take the form of (a) a review of the evaluation report by an external peer group and/or (b) direct participation of external consultants in the conduct of the evaluation.

Figure 3: Schematic Auto-Evaluation Process

Steps	Tasks	Technical Staff Involvement	PBEE Involvement
AE planning and timing	▪ Selection of PEs/clusters to be evaluated. ▪ Nomination of AE Managers. ▪ TORs preparation: selection of issues to be covered, methodology, external inputs (consultants, etc.) and resources required.	▪ The ADG selects PEs/clusters to be evaluated in consultation with Division Directors, and approves the selection of AE Managers (usually Service Chiefs or PE Manager for single PE evaluations, and Division Directors for clusters). ▪ Evaluation issues must include the most pressing questions to which the concerned staff would like to find answers, collected for instance through a brainstorming meeting. ▪ TORs should be circulated to all concerned staff for comments.	▪ Reviews AE proposals and provide advice. ▪ Can provide *ad hoc* assistance in TORs preparation (issues, methodology). ▪ TORs submitted to PBEE for review and clearance.
Administrative / managerial set up	▪ Allocate evaluation-related work to staff, under the oversight of the AE Manager. ▪ Hire consultants if needed.	▪ Depending on the size of the evaluation, the AE Manager may wish to involve other staff for desk reviews, preparation of questionnaires, data analysis, etc.	▪ Can search for consultants and propose them to AE Managers.
Conduct of the evaluation	▪ Collection, collation and analysis of evaluation material by the evaluators (AE manager, PE staff and/or consultants).	▪ Staff at Headquarters and in Decentralized Offices are a key source of information and perceptions for AEs, to be collected through individual interviews, focus groups and/or email.	▪ Can provide methodological assistance (e.g. in surveys design and analysis).
First draft of the evaluation report	▪ Elaboration of conclusions and recommendations. ▪ Report preparation. ▪ Report submitted to ADG	▪ Draft discussed with, and commented upon by, all concerned staff including concerned Regional Offices.	▪ Is copied and comments upon the draft report.
Peer review (where part of the process)	▪ External review of the draft report by knowledgeable peers.	▪ Discuss with peer review panel.	
Preparation of the final report	▪ Overseen by AE manager, based on comments received and ADG review.	▪ All concerned staff and Decentralized Offices receive a copy of the final report, and their attention is drawn to those recommendations that concern them.	▪ Collects all final reports. ▪ Presents summaries of AE reports to Governing Bodies.

The main rationale behind external inputs is to combine the advantage of auto-evaluation in terms of ownership of the results with objectivity and independent judgement. The use of external inputs will be especially important for:

- High-priority, visible area of work, receiving significant funding and reaching an important or strategic audience (e.g. CODEX, FIVIMS);
- Programme entities considered for cancellation or extensive re-formulation; and
- Programme entities and clustered auto-evaluations involving several Divisions or Services, so as to maintain neutrality.

As external inputs normally have a cost, they may be brought to bear more readily and cost-effectively in the case of well-funded AEs, i.e. AEs that cluster several programme entities or that pertain to large, well-endowed PEs.

3. Roles of PBEE

The Evaluation Service (PBEE) will provide *ad hoc* assistance in the preparation of TORs; methodological advice to divisions and PAIA chairpersons through guidelines and training on the design and conduct of evaluations; on-demand methodological support in the design of questionnaire surveys; facilitation for brainstorming meetings and SWOT sessions; help in finding consultants; and comments on the draft report.

Auto-evaluation TORs should be reviewed and cleared by PBEE, as already the case for the TORs of field project evaluations. PBEE will also prepare summaries of completed AEs as a basis for periodic reporting to the Governing Bodies on auto-evaluation results.

The quality of TORs, together with the quality of the auto-evaluation report, will form the basis upon which the Evaluation Service will allocate financial support to auto-evaluations during the initial introduction of the modality in 2003-04.

PART III: PLANNING FOR AUTO EVALUATION

A. Defining the Evaluation Issues

The first and in many ways the most critical step when planning an evaluation is to define a set of issues to be evaluated. Asking the right questions is absolutely crucial if one wants to find useful answers. Once collated, the list of issues will be used to prepare terms of reference, questionnaires, checklists for interviews, etc. On each issue, the evaluators will be required to present their conclusions and, if needed, their recommendations.

A basic common set of evaluation criteria against which to evaluate programme entities is required to ensure that all AEs yield more-or-less comparable results. The common criteria for AEs are defined as follows:

a) conformity to the Organization's mandate; relevance to the strategic objectives and use made of FAO comparative advantages;

b) relevance to the needs of countries, international community and other target users of FAO services;

c) quality, coherence and clarity of programme entity's design, including for cluster AEs coherence between related PEs;

d) strength and use made of internal and external partnerships;

e) adequacy and management of staff and financial resources, cost-efficiency;

f) overall performance in output production, particularly against qualitative and quantitative targets set in the MTP;[5]

g) quality and adequacy of outputs produced as assessed by subject-area specialists and/or by their actual or potential users;

h) effectiveness (including cost-effectiveness) in the realization of outcomes and achievements at the programme entity objective level,

i) contribution to PAIAs and the Gender Plan of Action, particularly when such contributions were planned in the MTP; and

j) the extent to which the benefits and improvements realized are likely to be sustained in future.

However, limiting AEs only to this list may reduce their usefulness in terms of bringing improvements to programme design and implementation. So AE managers will have to enrich and flesh out the common issues outlined above, so as to make them more specific to their own programme.

The starting point to flesh out the evaluation issues is the programme entity as planned, including rationale, objective(s), outcomes, outputs and indicators. Evaluations routinely examine whether the objectives of the evaluated intervention are realistic and appropriate, as well as how they are being met.

[5] Taking into account the difference between real and projected resource levels, as appropriate.

In addition, auto-evaluations may need to study emerging issues, such as unforeseen problems or opportunities for future action.

The best way to broaden the list of evaluation issues is through a participatory process reviewing the expectations of concerned staff and implementing partners, for instance in a series of AE planning meetings. Canvassing participating staff's expectations will help ensure they view the evaluation as relevant and useful. All consulted staff may, for instance, be asked the same two "triggering questions":

- What do you want to learn from, or illustrate in, this evaluation?
- What problems do you think should be tackled in the recommendations?

One danger of broad consultations is that they often yield long lists of every conceivable question that might be asked with respect to the general topic of concern. Information, however, does not come cheap; it takes time, effort and resources to collect reliable data and come to sensible conclusions and recommendations. Long lists of issues may also result in annoyance and frustration on the part of many respondents. As a rule of thumb, the list of issues for an auto-evaluation should *never* exceed two pages.

In the process of narrowing the list of issues, one should attempt to sort them out in sets of "connected" issues, so as to reveal similarities and overlaps between them. This technique is of great help to determine the most synthetic formulation for questions to be addressed by the evaluation.

When assessing the respective priority of various evaluation questions, one should always consider the possible use of the information collected under a particular issue. What consequences will a particular response to the question at hand have on the programme entity being evaluated? Does it matter that we know more on this issue to define what to do next in this programme entity?

Note that evaluations are not exactly free of conflict. On the contrary, they tend to generate friction and arguments, inasmuch as they are perceived as key to the future of the evaluated programme. Defining the issues to be evaluated in a participatory fashion may lead to some resistance and unrest. One way to tackle this is simply to drop contentious issues, but a more productive approach is to try and ensure that contentious issues are tackled in a neutral and independent fashion during the evaluation, for instance by bringing to bear external inputs (peer review panels and/or consultants) preferably on contentious issues.

Figure 4: Examples of Issues for PE Auto-Evaluations

The following are only **examples** of what a list of evaluation issues can look like, **not a limitative or prescriptive list.** The list is probably too generic to work well with any particular evaluation. The categorization in design issues, output issues, etc. is likewise only one way to structure evaluation issues. The right way to elaborate a list of evaluation issues is through a participatory process involving an array of programme staff and partners. All auto-evaluations should, however, include a description and assessment of achieved outcomes.

Design Issues
▪ Are the PE strategy and hierarchy of objectives (outputs, outcomes, objective) coherent & achievable?
▪ Are the resources allocated to the PE sufficient to deliver the outputs?
▪ Does FAO have a clear comparative advantage, mandate and priority on the subject matter?
Implementation and Process Issues
▪ Are planned resources (human and financial) actually available to the PE and well utilized?
▪ Is FAO working with the right partners and competencies on the subject matter?
▪ Are outputs produced at a reasonable cost and with accepted quality standards?
Output Issues
▪ Which outputs did the PE produce during the evaluated period, and how does this list compare with planned outputs?
▪ How do biennial outputs contribute to their major outputs?
▪ Is there an effective dissemination strategy for FAO outputs?
Outcome Issues
▪ What is the actual audience? What sort & number of users are reached by FAO products & services?
▪ What do they think of FAO outputs & what do they do with them?
▪ Are there any unplanned outcomes (positive or negative) resulting from the PE?
Objective-level Issues
▪ What contributions to improved decision-making at international, national or sub-national levels can be documented or conjectured from existing evidence (e.g. in governments, donors, UN agencies, community organizations and NGOs)?
Cross-Sectoral Issues
▪ How has the evaluated initiative contributed to the goals of the PAIAs it was planned to participate in?
▪ Was any progress made on gender mainstreaming, e.g. in the priority areas identified in the Gender Plan of Action (gender-segregated data, gender-sensitive communication strategies, equal access to natural resources and agricultural support systems, gender-sensitive policy and planning, etc.).

B. Deciding on Evaluation Methodology – Information Sources and Techniques

Once a list of evaluation questions has been agreed to, the next step is to identify possible sources for the answers. This is usually referred to as the evaluation methodology. The methodology is composed of a set of *sources* of information and *techniques* to extract this information. The main sources of information in AEs will be:

- **FAO staff** involved in the programme entity or PAIA to be evaluated;
- **staff in partner organizations**, who have been associated with the evaluated work;
- **primary users**, who have used or are supposed to use FAO outputs.

As explained in Part II, the "data collection difficulty" rises as one moves along the user chain and further away from FAO. Surveying or interviewing FAO staff and partner organizations is easier than surveying primary users. Secondary users are even more difficult to access than primary users, and surveying end-beneficiaries at the grass-roots level would appear impossible, at least if one wants to relate their perceptions to the FAO Regular Programme.

Some of the most common evaluation techniques for auto-evaluations are listed below. Part V: Auto-Evaluation Techniques provides advice on how to use each of these techniques. Here we are chiefly concerned about how to select and combine them when planning for an AE:

- **Indicators** have been specified in the Medium-Term Plan. Their assessment provides much of the basic information for evaluation;
- **Desk studies and annotated bibliographies** can be very useful to review the documentation accumulated by the evaluated programme (back-to-office reports, meeting minutes, and of course programme outputs themselves, since they are often of a documentary nature);
- **SWOT analysis** has been applied extensively by private and public organizations to help identify strengths, weaknesses, opportunities and threats in a qualitative way;
- **Questionnaire surveys** of staff, partner organizations and primary users are likely to be of help in many auto-evaluations, in particular to assess strengths and weaknesses in a quantitative way, i.e. by estimating the percentage of users satisfied with a given FAO product or service. They can also be used to collect observations and recommendations through textual inputs;
- **Semi-structured interviews** of key staff, partner organizations and primary users, either individually or in groups (**focus groups**), may provide richer and deeper material than questionnaire surveys do. Semi-structured means that the interviewer follows a checklist of issues but is free to explore issues beyond the list. Such interviews are most useful when one wants to collect qualitative information, such as "client stories" (narratives of how a particular user has applied or disseminated FAO's products and services);
- **Country case studies**, involving travel to a sample of countries to analyze programme outcomes, are costly exercises that are particularly useful if the evaluated programme has achieved quite a lot in a limited number of countries;

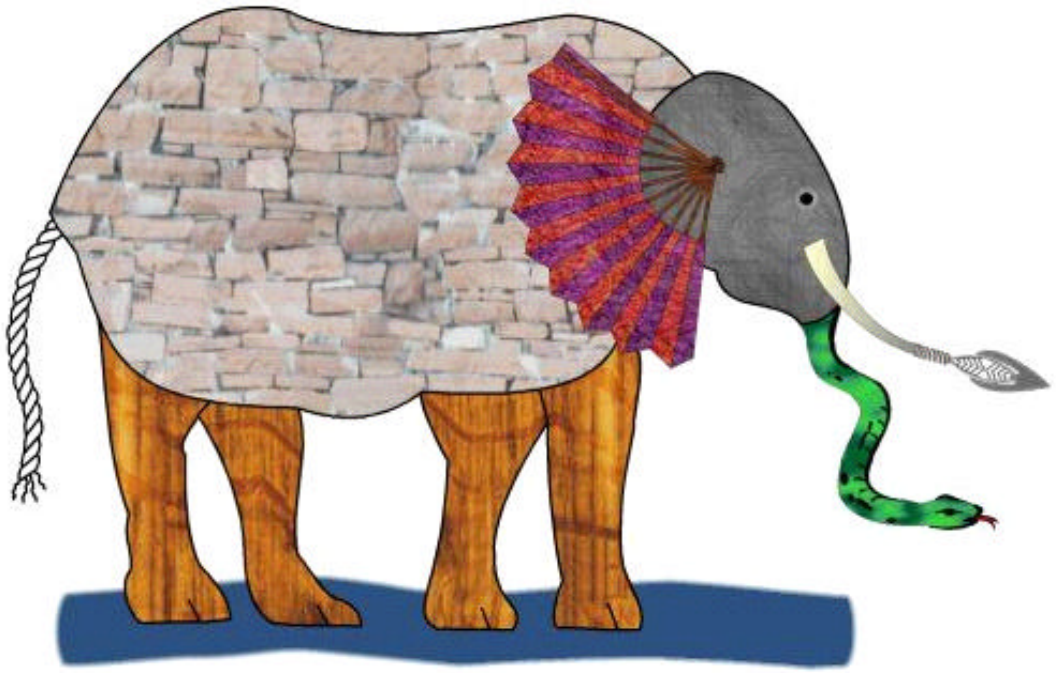
Figure 5: Advantages and drawbacks of some evaluation techniques

Evaluation Techniques	Advantages	Drawbacks
Indicators	▪ Strong accountability tool, since determined at planning stage. ▪ Measure changes and trends. ▪ Can be qualitative (short statements, examples…).	▪ Often difficult to verify (especially objective-level indicators). ▪ Fail to capture unexpected developments. ▪ Describe but do not explain.
Desk studies / annotated bibliographies	▪ Good starting point for an evaluation. ▪ Reduce the time needed for consultants or staff to access programme details or status of research.	▪ May take time to assemble. ▪ Usually undertaken by junior staff but need good supervision, or miss essential points.
SWOT analysis	▪ Participatory and transparent. ▪ Good strategic tool: help focus on what is important and help define recommendations.	▪ Time consuming: many staff involved. ▪ Some people regard such facilitation techniques as childish.
Semi-structured individual interviews	▪ Help capture complex programmes aiming at varied outcomes. ▪ Capture processes and problems. ▪ Help understand the meaning of a programme to its stakeholders.	▪ Take time to interview and analyze, cannot be automated. ▪ Not good for programmes with repetitive, predetermined outcomes. ▪ Some audiences distrust qualitative research, prefer statistics.
Focus group interviews (usually semi-structured)	▪ Same advantages as individual interviews. ▪ Work by consensus between informants. ▪ Collect views from a good number of informants. ▪ Quick to identify important issues.	▪ Same drawbacks as individual interviews. ▪ May inhibit the expression of minority views.
Questionnaire surveys	▪ Powerful data collection tools: larger group of informants. ▪ Objectivity: data collected in a standard and formal way.	▪ Designing good questionnaires is difficult. ▪ Response rate often low. ▪ Often no obligation to respond to questionnaire, hence the sample is biased towards the most opinionated.
Web statistics	▪ Cheap (GILW does the statistics for you). ▪ Crude analysis of the volume and geographical origin of audience, and of documents / pages most downloaded.	▪ Difficult to interpret, need to eliminate hits by search engine “robots”, etc. ▪ Misses vital data: gender, nationality, occupation. ▪ Geographic data biased towards developed countries access providers. ▪ Quality of visits is often more important than quantity.
Country case studies	▪ Best way to capture rich results at the country level.	▪ Need good planning and administration. ▪ Costly. Need strong supervision of consultants (where used).
Expert panels	▪ Provide accountability and transparency. ▪ Help confirm evaluation validity.	▪ Can cost from US$5,000 to US$15,000. ▪ Can produce conflict if act as parallel evaluation teams.

- **Web statistics** as collected by GILW can give an idea of the quantity and geographical origin of those consulting a particular website; and
- **Expert panels** have been used extensively by PBEE as an independent quality control process for its own evaluations – i.e. to review evaluation reports and make sure that they conform to professional standards – and as a way to incorporate a broader perspective into the evaluation.

It should be stressed that there is no such thing as a perfect evaluation technique. They all have their advantages and disadvantages, biases and cost range, which is why they should be carefully selected based on the type of question to be answered. It is common practice, particularly on the most important issues, to collect information from different sources so as to put together a less biased representation of reality. This is called triangulation, a very useful but tricky evaluation practice.

Figure 6: Triangulation, or the art of combining different perceptions into one representation.



In a famous Indian legend, six blind men are feeling an elephant for the first time and are imagining it in their mind. One believes the legs are tree trunks, another one describes the tail as a rope, the body is felt as a wall, the trunk as a snake, the ears as fans, and the tusks as spears. None of the blind men get it right on his own and they all disagree on what it is there are faced with, yet one could still put together all the pieces and draw a picture loosely resembling the animal.

Pictures by Paul Galdone, in "The Blind Men and the Elephant; John Godfrey Saxe's version of the famous Indian legend", Whittlesey House, 1963. Full poem in http://www.noogenesis.com/pineapple/blind_men_elephant.html

A combination of quantitative and qualitative techniques, as well as an awareness of common biases in evaluation techniques, are key ingredients in a successful triangulation, as is an analysis of the respondents' "stake" in the intervention being evaluated. For instance, people tend to be rather positive about their own achievements, quite understandably. Recipients and close partners of a programme often display a positive bias as well, while UN organizations not associated with FAO may suffer from a slight negative bias. The views of competing organizations are usually very slanted and should be checked systematically.

C. Estimating a Budget

The next step in planning for an evaluation is to estimate how much resources are going to be required. Note that this is an iterative process. If the budget required for the selected evaluation methodology exceeds available resources, one would have to revise the methodology, deleting the most costly data collection techniques or decreasing consultancy time, etc. so as to fit in available resources.

Outlining the methodology in sufficient detail makes it easier to price each evaluation technique (surveys, field trips, etc.) with reasonable precision. A key element of the budget for an auto-evaluation will be the required external inputs, in terms of consultant fees and DSA, DSA for peer-review groups (usually not remunerated otherwise) and related travel costs. The cost of such external inputs may be covered in part by PBEE during an initial period of two years (2003-2004).

Internal inputs – mainly staff time for the AE Manager and others involved – should also be estimated and priced in the auto-evaluation budget as part of PWB planning. As an order of magnitude, the total cost for auto-evaluations (including both external and internal inputs i.e. staff time) will normally range from US$20,000 and US$50,000 depending on the size of the evaluated PE(s), or about 2 to 3 percent of the evaluated PE(s) budget.

Even though the concerned divisions may not decide to include a peer group review and one or several consultants in all their auto-evaluations, the management cost in terms of staff time is not negligible. In general, one should be cautious about under-estimating the resources needed for a good-quality evaluation.

Departments and divisions may wish to combine several programme entities into one auto-evaluation. Such "cluster evaluations" regrouping *related* PEs may in many cases be more cost-effective than reviewing each programme entity in isolation.

D. Drafting and Circulating Terms of Reference

By this stage, one should have all the elements for drafting TORs: a short background section, a list of issues to be evaluated, a tentative methodology, a description of the persons involved in the evaluation (respondents, FAO staff and consultants managing the process and writing the report), and a budget. An indicative TORs outline is provided in Annex 1.

TORs need to be circulated to all concerned staff in order to verify that a consensus was reached on the issues to be reviewed and the broad thrust of the evaluation process. In general, the shorter the TORs, the better. A few pages normally suffice.

TORs should be cleared by the concerned ADG and by PBEE.

PART IV: MANAGING AN AUTO EVALUATION

A. Sequencing Data Collection Techniques

As seen above, it is common practice to combine different sources of information during an evaluation. The next section (Part V: Auto-Evaluation Techniques) provides guidance on individual evaluation techniques. But how is one supposed to phase the selected evaluation techniques so as to fulfill one's evaluation goals in the least possible time? While there is no standard way to sequence evaluation techniques, it is possible to give some advice.

A desk study of printed and electronic material is a good way to start. It may roughly establish what the original plan was[6], draw a list of all outputs produced over the evaluated period, shed some light on implementation constraints and identify the main implementation partners. It may also provide a few pointers on achieved outcomes, which may be alluded to or even described in some detail in back-to-office reports. A desk study can also include a review of the GILW web statistics, where appropriate.

If the PE (or cluster of PEs) works with a significant number of implementation partners and/or with a large group of well-identified "users", it may be a good idea to use questionnaires to collect their perception of the evaluated PE(s) and their use of its outputs. If such an approach is adopted, the questionnaires should be drafted early on, based on evaluation issues identified in the TORs. Allow some time – a minimum of two to three weeks – for the development, circulation among concerned staff, testing, finalization and dispatching of questionnaires. Respondents will need about a month to fill in the questionnaires and mail them back. Reminder letters or emails are usually necessary to elicit good response rates.

If semi-structured interviews of staff, partners and clients are envisaged, try and identify the persons to be interviewed as early as possible so as to fix appointments in advance. Informants are rarely available when evaluators are ready to interview them. Since the evaluator is requesting the meeting, she/he will have to fit the schedule of her/his informants rather than the contrary. The usual practice is to interview the project or programme staff first, and then move on to external informants. This allows the evaluators to verify with users and partners the outcome claims originating from the staff.

The same rules apply to country case studies and field trips, which need to be planned well in advance so as to ensure the availability of key stakeholders in the country.

Note that a questionnaire survey can be combined with follow-up semi-structured interviews, e.g. by phone or email, for those questionnaire respondents that appear particularly knowledgeable. Conversely, semi-structured interviews can help prepare a questionnaire survey by identifying issues and typical answers to evaluation questions.

Staff are a primary source of information to verify and describe in greater detail the reported outputs that may have been identified through a desk study, and to identify

[6] Extracts from PIRES MTP and PWB applications will provide a good picture of the original intentions, resource allocations and planned partnerships.

implementation constraints and opportunities. Staff interviews also help collect more information on outcomes and identify partners and primary users more precisely than is usually done in implementation reports.

Establishing with the concerned staff a list of strengths, constraints, threats and opportunities through a SWOT exercise is a good way to establish recommendations that the project staff would be keen to implement.

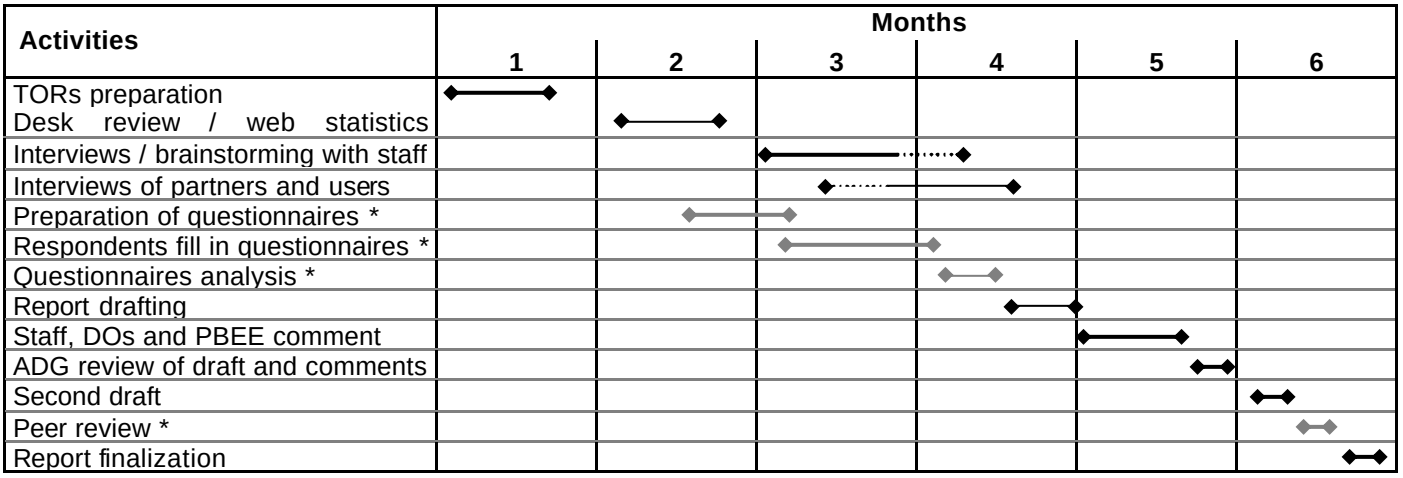
Once most concerned staff have been interviewed, the evaluator should move on to interview a wide selection of partners and primary users. These persons represent one of the prime knowledge repositories on the quality of the evaluated PE(s) outputs and the sort of uses the PE(s) outputs are put to, i.e. outcomes. Ultimately, external informants that are closely involved with the evaluated PE(s) are ideally placed to pass a judgment on the PE(s) relevance and usefulness.

Some primary users may be in a position to narrate a few achievements at the PE objective level, i.e. significant changes in decision making or policy within governments, donors, or field projects that have been influenced by the evaluated FAO initiative.

If this is the case, one could envisage contacting the concerned "secondary users" (i.e. decision makers) to further describe any achievement at the PE objective level. For reasons explained in Part I, however, such achievements at the PE objective level are difficult to identify, and generally not the result of the FAO RP alone. Besides, decision makers tend to be quite busy and unaware of the influence FAO may have had on the decisions they have taken. So interviewing them will not always be useful.

Once staff, partners and users have been interviewed or surveyed, the AE manager should lay down her/his observations, conclusions and recommendations in a draft report to be circulated to the concerned staff, Decentralized Offices and PBEE. As already explained in Part II: Procedures, the concerned ADG will then review both the report and comments, and will indicate to the AE manager how comments should be reflected in the second draft.

Figure 7: Example of an Auto-Evaluation Timeline



* Questionnaire surveys and peer reviews are shown in grey because they are optional, to be brought to bear preferably on the most successful or strategic programme entities.

As a final step in the auto-evaluation process, a peer group could be assembled, either physically (recommended) or virtually (through emails or message boards) to review the consolidated report and comment upon it. Peer group reviews are recommended for the auto-evaluation of high-priority, visible and well-funded area of work, for cross-divisional PEs or cluster evaluations, and for programme entities considered for extensive re-formulation or cancellation.

B. Tips for Report Presentation

An indicative report outline is proposed in Annex 2. The easiest way to structure an evaluation report is often to follow the list of issues in the TORs. This will ensure that the reader can compare the AE report with its TORs. Obviously, this approach only works inasmuch as the TORs issues are well ordered and avoid repetitions.

A table presenting the main quantitative indicators collected during the evaluation can be useful to summarize the programme entity achievements from a quantitative standpoint.

Detailed client stories are a very good tool to present the best (or worst) that a particular programme has to offer in a lively, telling manner. They can also evidence how, i.e. by which causality mechanisms, a project works or does not work.

Signed quotes, displayed prominently in the final report e.g. within text boxes, will usually convey a more powerful message than anonymous quotes. It is a good practice to identify among your returned questionnaire and interview records the verbatim quotes you want to use in the report, and then ask for the consent of their authors to quote them by name.

Evaluation conclusions involve subjective judgment calls, e.g. on whether or not specific activities are worth the expense and efforts invested in them. The AE manager or report writer should defend her/his opinions, but also identify them clearly and distinguish them from more factual descriptions through the use of phrases such as “the evaluators conclude that…” or “we are of the opinion that…”.

If a team of evaluators is involved, all of them should participate in report writing. The evaluators should seek consensus and agree on their findings, conclusions and recommendations. If strong disagreements surface within the evaluation team and cannot be resolved, those holding the minority view may consider writing a dissenting opinion that will be annexed to the main report. This is an extreme and seldom-used measure; usually evaluators manage to minimize their differences.

Recommendations are the most important part of the report, one that will determine if the evaluation is relevant and useful. Recommendations should be few in number (typically from 5 to a maximum of 15), realistic and feasible, important to the success of the enterprise, creative and imaginative, and actor-specific.

Any proposal for a new or revised programme entity should provide a rough sketch of the PE design, in order that this can be used as an input in the next MTP preparation.

There should always be a set of recommendations which do not require increased resources, in addition to any which do call for additional resources. Throwing money at problems is always a tempting option and sometimes a specific intervention does require more money or staff to achieve its objectives. However, FAO has to operate within limited resources.

Once the first draft has been produced, it should be reviewed and commented upon by the staff of the evaluated programme entity(ies), other concerned units including Regional Offices and PBEE. Following the ADG review of the report and comments, the AE manager will oversee the preparation of the final report or, if a peer review has been opted for, of a second draft to be reviewed by the peer group.

Note that all comments cannot usually be incorporated in the final report of an evaluation. Some of them may be incidental to the evaluation. Others may come in contradiction to the evaluators' argumentation. While it is a good practice to discuss or describe a broad array of ideas and perceptions about a given programme or project, the evaluators need to draw their own final conclusions and recommendations.

C. Quality Standards for Auto-Evaluations

Systematic auto-evaluation is a new process for the Organization and the initial years will constitute a learning period during which the methodologies and quality standards will be progressively refined.

Auto-evaluation is conceived as a rapid evaluation process, one that uses "quick-and-dirty" techniques to assess achievements and develop pertinent recommendations. Nonetheless, ensuring minimum quality standards is a key concern. How quick and how dirty should an auto-evaluation be?

The following criteria will be used to decide whether or not an AE meets minimum quality standards:

- The AE should make use of **external inputs** and must elicit **feedback from users, partners and peers**, though not necessarily in a scientific manner.[7]
- AE goes much beyond a passive description of activities. An AE exercise should review PE **design** and **relevance**, implementation **constraints** and **opportunities**, **outputs**, **outcomes** and **achievements against the PE objective,** though assessing the latter is recognized as a difficult task.
- Planned contribution to the **Gender Plan of Action** should be systematically reviewed and their realization assessed, as well as the contributions to other **PAIAs**.
- The report should clearly identify and assess **outputs and outcomes** from a quantitative and qualitative standpoint, making use of **MTP indicators.** Vague and woolly wording must not be used.
- The report need not exceed 30 pages, but it must display a **critical outlook** and include **precise, creative recommendations**.

AEs that fail in a significant manner to meet these criteria will not be eligible for PBEE funding during 2003 and 2004.

[7] In other words, it is recognized that surveys conducted within the framework of auto-evaluation may be based on non-random and possibly biased samples.

PART V: AUTO - EVALUATION TECHNIQUES

A. Reviewing and Defining Indicators

As explained in the previous section, MTP performance indicators at the level of outcomes and programme entity objective will constitute an important building block for the list of issues to be explored in any auto-evaluation. Auto-evaluations will have to collect data against each MTP indicator so as to verify progress against pre-set targets, if available and taking into consideration any discrepancy between resources as planned in the MTP and as actually made available to the evaluated PE(s).

Whenever possible – i.e. when the programme entity structure remained largely unchanged from the MTP 2002-2007 to the MTP for 2004-2009 for the work to be evaluated – it is recommended to use the more elaborate version of those indicators as defined for the MTP 2004-2009 and as displayed in the PIRES interface.

The MTP indicators are not a limitative list. The auto-evaluation managers may wish to collect additional or alternative indicators, or they may during the course of their evaluation collect interesting quantitative data shedding light on the advancement of their programme entities.

The New Programme Model defines indicators as variables (e.g. "number of countries where such and such benefits are produced", rather than "such and such benefits produced in 5 countries") that can be objectively collected and verified through a clear methodology, to document or describe to what extent the PE objective or outcomes are being met.

Figure 8: Indicators, targets and achievements

Indicators	Targets	Achievements		
		2001	2002	2003
List of countries collecting data using methodology XXX	20	5	10	15
Number of national staff exposed to new approach through workshop	2,000	500	1,800	2,500
Examples of lessons learned and agreed at international conference	Consensus about issue YYY	Consensus achieved	n.a.	n.a.

Defining indicators as variables allows comparisons between various values taken by the variable at different dates, or between targets and actual achievements (Figure 8). The best way to use indicators is actually through their repeated assessment over a period of time (e.g. baseline, mid-term, end of project) so as to demonstrate or detect trends over time. This may not always have been done systematically in the past, but evaluators should seek to form an understanding of the situation at the beginning of a Technical Project (TP) or 5-6 years ago for a Continuing Programme (CP).

Care should be taken not to adopt long lists of indicators, since listing them is much easier than actually measuring them with reasonable precision. However long a list of indicators, it cannot describe everything that may be said or measured about a given work.

Finally, indicators only indicate change. They do not explain why change is happening.

B. Desk Studies and Literature Surveys

A desk study is a very good way to start an evaluation. It usually entails a compilation and summary of all relevant documents, such as programme documents, workplans, back-to-office reports, meeting minutes, and of course programme outputs themselves since they are usually of a documentary nature in the case of the Regular Programme.

Extracts from PIRES MTP and PWB applications will provide a good picture of the original intentions and of the progression of usually declining resource allocations, from the MTP stage to the PWB. Planned partnerships can also be extracted from PIRES.

Desk studies are often performed by junior staff or consultants. This will allow them to familiarize themselves with the "nuts and bolts" of programme design and implementation, and will reduce the amount of time more senior consultants or staff have to spend on understanding the programme details.

Literature surveys are somewhat similar to desk studies in that they entail the reading, summarizing and sieving through of a vast documentation. The difference is that literature reviews are performed on documents external to a programme, i.e. publications (either in print or electronic) produced by the technical or mass media, by other programmes or organizations. Such reviews may prove useful in various ways:

- to describe the main issues and opinions being debated by "opinion makers" with whom FAO interacts (or wishes to interact);
- to see if FAO is making a unique contribution or is merely repeating the work of others; and
- to collect references to FAO work in publications of others, so as to gauge the FAO audience.

Some publications that used to be available only in print are now posted on Internet. Web searching technologies have made the review of at least the main media outlets much easier than in the past. A search for a few, well-chosen words or group of words from a FAO publication may often yield valuable hints on which other publications or media are quoting FAO. Most technical publications, however, are not accessible through the Web and reviewing those remains therefore quite time-consuming.

C. SWOT Analysis

SWOT analysis is a participatory methodology by which stakeholders of a specific project or staff in an organization identify their or their unit's strengths and weaknesses, as well as the opportunities and threats they face. It is normally carried out with the help of an external facilitator.

In the context of AE, such kinds of facilitated exercises can be envisaged within the major units involved in the management of a PE at Headquarters or in Regional Offices. As this

is quite time-consuming, decisions have to be made as to when the insights gained are likely to be sufficiently valuable.

Note that some people regard such facilitated processes as childish and may prefer more conventional meetings with someone taking notes and then presenting a draft for further discussion.

The first step in a SWOT session is to ask participants to fill in cards with the strengths, weaknesses, opportunities and threats they perceive as important. These are then placed on corresponding pin-boards. Each idea should then be discussed and agreed upon by the whole group. Only those cards on which the whole group agrees would normally remain on the pin-boards, as a reflection of group consensus. Duplicates or very similar ideas should be rephrased as one card. Alternatively, the facilitator may write ideas up on flipcharts.

Definitions:

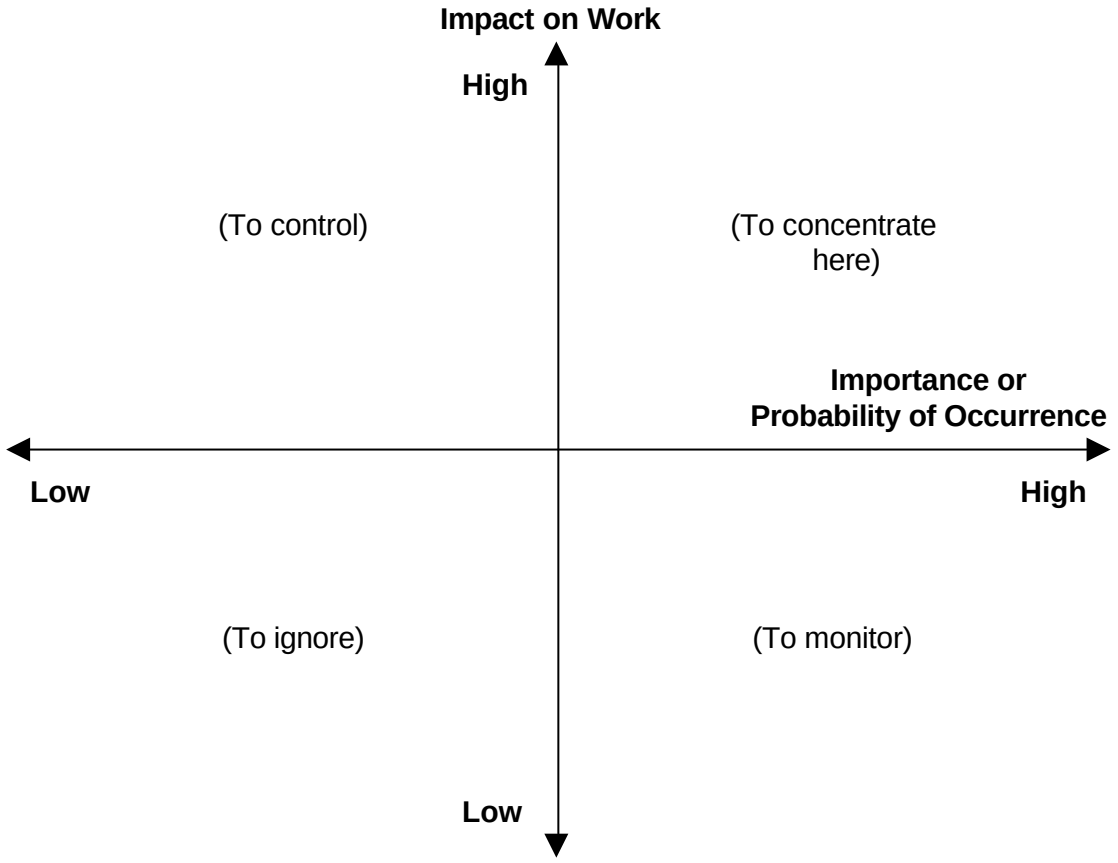
- **Strengths**: qualities, assets or strong points inherent to or associated with the group, which may enable the group to fulfill a useful and/or expanded role.
- **Weaknesses**: traits that have a direct negative impact on the group's work performance and hinder its ability to perform an appreciated role and remain relevant.
- **Opportunities**: external factors, circumstances or trends which can potentially and favourably affect a group's ability to operate, and can lead to a positive development of the group's role if they are exploited.
- **Threats**: external circumstances, competition or risks which might unfavorably influence a group's activities or relevance, and/or might result in an undesirable development.

A second and often omitted step, once a list of strengths, weaknesses, as well as the opportunities and threats has been agreed to, is to give scores (e.g. on a scale from 0 to 10) to each card along the following two dimensions: 1) actual or likely impact on the work at hand; and 2) intensity (for strengths and weaknesses) or likelihood (for opportunities and threats). Once the scores have been agreed to, the cards can be placed on a matrix as displayed in Figure 9. The matrix helps focus on what are the strengths, weaknesses, opportunities and threats that are perceived as really important by the group, i.e. those located in the upper left corner.

Well-crafted SWOT items are quite instrumental in identifying recommendations, since each category has distinct implications at the strategic planning level. They call for different responses:

- **Strengths** can be used and built upon. They will help exploit the opportunities of the environment or overcome threats and potential difficult situations;
- **Weaknesses**, to the extent possible, should be eliminated or minimized. They hamper the effective exploitation of existing or future opportunities of the environment and erode the ability to react to difficult situations;
- **Opportunities**, if not taken advantage of, are worthless. They only matter if one decides to exploit them or has the potential to do so; and
- **Threats** have to be overcome or mitigated.

Figure 9: The S.W.O.T. Matrix



D. Semi-Structured Interviews

Semi-structured interviews use a checklist of open-ended questions, but the interviewer is free to go over and beyond the list. Responses to questions provide the evaluators with quotations, which compose the raw data. Quotations reveal the respondents' levels of emotion, the way in which they view the situation at hand and of their role therein, their thoughts about what is evaluated, their experiences and perceptions. Semi-structured interviews are most useful when one wants to collect complex, qualitative information, such as perceptions or client stories, i.e. narratives of how a specific user has applied or disseminated (or not) FAO's products and services.

Qualitative interviews may be used as an exploratory step before designing more quantitative, structured questionnaires to help determine the appropriate questions and categories. Conversely, interviews may be used after results of more standardized measures are analyzed to gain insight into interesting or unexpected findings. Their value as compared with questionnaire surveys is that they can probe respondents and get a much clearer and richer picture of their views than questionnaires can.

Quantitative surveys are usually given more credit than qualitative research, even though the former can be dismissed on methodological grounds by those who disagree with the

findings. Qualitative research is most potent and convincing when the report highlights selected quotations, i.e. the actual words of surveyed staff, partners or clients, so as to convey their perceptions and emotions in an unadulterated manner.

E. Focus Group Interviews

A variation of the semi-structured interview is the focus group interview, in which informants of a specific type or stakeholder group are interviewed all at once. They have more or less the same advantages as individual interviews. In addition they can rapidly collect views from a broader array of users, partners or staff than individual interviews. The downside is that they usually inhibit the expression of minority views among the group.

In the context of auto-evaluation, focus group interviews could prove useful as a way to collect stakeholders' views during country case studies, for instance if the evaluated intervention has worked with a broad array of local partners and NGOs.

A focus group can be defined as a group of interacting individuals having some common characteristics, brought together by a facilitator who uses the group and its interaction as a way to gain information about a specific issue. A focus group is typically composed of 7 to 10 people. If one wishes to interview a larger group, the normal practice is to convene several focus groups.

One of the advantages of focus groups is that the group can check the validity of individual ideas. Focus groups are also a great way to study interactions, consensus or disagreements between participants. However, they are not appropriate for collecting individual informants' perceptions.

They provide data more quickly and at lower cost than individual interviews, and are easier to organize than questionnaire surveys. But they require skilled facilitators. Here are a few characteristics of a good facilitator:

- able to keep the discussion flowing and on track;
- courteous and friendly, yet capable of probing people and asking dominant participants to let others talk;
- sensitive to the mood of the group, capable of cheering it up and encouraging shy participants to talk;
- self-controlled, avoids influencing the group, not giving clues about "desirable answers";
- good listener, able to spot subtle allusions and veiled criticism;
- external to the reviewed programme so as not to be perceived as partial; yet
- knowledgeable about the evaluated work and able to tell important ideas from incidental ones.

The choice of group participants is crucial to the group dynamic. Participants should obviously be knowledgeable about the specific issues to be explored. As a rule, try and interview each stakeholder group separately: programme staff should never be mixed with partners and beneficiaries.

Focus groups should not be rushed. The facilitator should start slow, establish a contact, describe the purpose of the evaluation and the ground rules for the session before moving to the first question. She/he should give people some time to think and pause for a few seconds after a participant has finished talking in order to let others jump in. The number of questions should be limited to 5 or 6. Questions should be open-ended and should be arranged in a logical sequence. At the end of each session, it is a good practice to summarize what was said to the participants and ask whether they have anything to add.

F. Questionnaire Surveys

Questionnaire surveys constitute powerful tools for collecting quantitative and qualitative information on programme performance. Designing and conducting questionnaire surveys is however far from easy, as anyone who ever tried to write a concise and user-friendly questionnaire can testify. Obtaining responses to questionnaires is often a challenge, and analyzing the responses has its own pitfalls. Low response rate is the biggest worry of surveyors using "self-completed questionnaires"[8], and will likely constitute a significant constraint in surveys launched within the framework of auto-evaluation.

In general, we can recommend using a questionnaire in auto-evaluation to elicit from a significant group of users, staff or partners and in a standard way their subjective judgments, attitudes, opinions or feelings about the usefulness of all or part of an existing project, such as the value of a work process or the quality, usefulness and actual use of outputs and services (i.e. outcomes).

A first step is to identify as precisely as possible the persons that may know best about the issues you want to get information on. Such a group of potential informants is called the survey's *target population*. General questionnaires arbitrarily sent to governments, donors, etc., should be avoided as much as possible, since they are more likely to annoy recipients than elicit purposeful responses. In auto-evaluation, the persons most directly in contact with FAO products and services would normally constitute the target population:

- Actual or potential primary users, i.e. those people, either inside or outside FAO, who access and use the services and products generated by the evaluated intervention, or could potentially use them but do not at the moment.
- Partner organizations outside FAO that provided help or have a stake in the project; and
- Programme staff participating in the evaluated PE(s).

Note that surveying secondary users – i.e. decision makers in governments, development agencies and NGOs/CSOs – would not be advisable in most AEs because decision makers often do not reply to surveys. Even if they would, they may not be able to trace back their decisions to FAO's advice, since they usually are not in direct contact with the Organization on the specific outputs being evaluated.

Once you have identified the target population, the next step is to draw a sample. This can be a complex procedure if one wants to make sure the sample is representative. In the case of auto-evaluation, we will probably deal with small samples, either because the target population itself is small, or because a large-scale survey would be too costly.

[8] Questionnaires sent to potential respondents for them to fill in at their leisure.

Figure 10: Advantages and disadvantages of some common surveying methods

Methods	Advantages	Disadvantages
Face-to-face Interviews	▪ Greatest ability to locate the target population and get a representative sample. ▪ Longer interviews are better tolerated. ▪ Interviewers can ask for clarification of responses.	▪ Cost per interview higher than for any other method, and extremely high for most AEs as target population is dispersed in member countries. ▪ Possible interviewer bias.
Telephone Interviews	▪ Fast. ▪ Less expensive than face-to-face. ▪ Interviewers can ask for clarification of responses. ▪ Long interviews are sometimes tolerated (less than face-to-face).	▪ Bad reputation due to resemblance to tele-marketing. ▪ Requires the phone numbers of a sample of respondents. ▪ Possible interviewer bias.
Feedback forms *	▪ Very fast. ▪ The training or workshop facilitator can motivate participants to respond, hence good representativity of results.	▪ Usually limited to the evaluation of one particular event, thus rarely applicable to entire PEs.
Mail Surveys	▪ Less expensive than interviews. ▪ Less intrusive than interviews. The respondents answer at their leisure.	▪ Quite slow to get the results (several weeks). ▪ Self-selected, and thus biased sample.
Email Surveys	▪ Very economical and fast. ▪ In developing countries, email access is often better than web access. ▪ Response rates usually higher (novelty of the method, no need to mail back the questionnaire).	▪ Many people dislike unsolicited email even more than unsolicited regular mail. ▪ Limited to target populations with good email connections. ▪ Self-selected, and thus biased sample.
Web Surveys	▪ Low data entry cost. ▪ Can automate skip patterns based on earlier answers. ▪ Response rates usually higher (novelty of the method, no need to mail back the questionnaire). ▪ Can be combined with an email invitation to take a Web survey.	▪ Limited to target populations with good internet access. ▪ Self-selected, and thus biased sample. ▪ Security must be well thought out to avoid anyone browsing that web page to answer. ▪ Identifying respondent may be a problem. ▪ Some may even answer several times.

*: E.g. questionnaire distributed at the end of a training or workshop.

Completing the questionnaire will in most cases be entirely up to the respondents, since in AEs, budget constraints rule out the use of enumerators to collect responses. The samples will therefore be "self-selected" rather than randomly selected. This type of sample typically generates biases and therefore forbids sophisticated statistical analysis. The busiest people or those who are generally happy with your work will not normally take time to fill-in a questionnaire. Instead you will end up surveying a few highly-motivated individuals who are either irritated about something your work did or very appreciative, the former being more frequent than the latter.

Self-completed questionnaires are an effective research strategy if one wants to identify the strengths and weaknesses of a programme as perceived by an outside audience. We should not assume, however, that the views expressed by self-extracted samples are representative of the target population in its entirety. We should take the feedback provided by such questionnaires with a "grain of salt", and consider that in our target population, only the most vocal people have responded.

Various methods can be used to dispatch the questionnaire to potential respondents. Figure 10 details the most common ones, together with their advantages and disadvantages. One critical factor to select the survey method will be how best you can get in contact with your sample: do you have their email address, street address or telephone number? Do they come and visit your web sites or attend FAO-sponsored workshops?

The next step is to translate selected evaluation issues into precise questions to put on the questionnaire. Most people have completed so many questionnaires in their life that even the least interested in surveys have some notion of how difficult it can be to design a good questionnaire. This is because questions are fixed. There is no possibility to include new questions on request from the respondent, and they cannot be explained in further detail if a respondent does not understand them (unless the questionnaire is read and filled in by a well-trained interviewer). Therefore the questions have to be well crafted to ensure that they will be understood correctly and answered.

The basic principle is to *KISS*… Keep It Short and Simple. If you present a 20-page questionnaire most potential respondents will give up in horror before even starting. Ask yourself what you will do with the information from each question. If you cannot give yourself a satisfactory answer, leave it out.

Questions can be classified as follows, depending on the sort of answers they call for:

- **Textual open-ended questions** – the respondent answers free text.
  Example: What aspects of the publication do you find the most useful? _____
- **Numerical open-ended questions** – similar to the above, but the respondent answers a number.
  Example: How many people read or browsed your copy of the publication? ____
- **Ordinal multiple-choice questions**, i.e. **rating scales** and agreement scales – the respondent is offered a sequence of options in a specific order.
  Example: do you agree with the following sentences?… agree / tend to agree / tend to disagree / disagree / don't know.
- **Nominal multiple-choice questions** – the respondent is offered a limited range of options in no specific order.
  Example: what section did you like most in the publication?: section 1 / section 2 / etc.

- **Ranking questions** – the respondent is asked to rank various items in order of importance.
  Example: Rank the following publication characteristics in decreasing order of importance: accuracy of information / usefulness of information / quantity of information / presentation and format.

Ranking questions tend to confuse respondents and are thus not advisable.

Multiple choice questions are easy to analyze and provide numerical data that are more convincing to some audiences than purely qualitative analyses. However, they cannot deepen or enrich an analysis as open-ended questions can.

Scales (i.e. ordinal multiple choice questions) are particularly widespread. They are easy to respond to and can be analyzed quickly. Their main advantage is that they allow comparisons between different features, provided a consistent scale is used. They are most often analyzed through percentages answering each option. Displaying the result of related questions using consistent scales side by side is often found useful, since the analyst can contrast the ratings of various features.

Open-ended questions explore the full range of views that your audience has and help you understand why people do or think as they do. They collect richer, more complex and detailed data and allow respondents to raise issues the surveyor may not have thought of. Professional surveyors, however, tend to dislike them because their analysis cannot be automated.

In practice, open-ended and multiple-choice questions are complementary and both types are used in most questionnaires. In particular, it is often a good idea to include a follow-up open-ended question after one or several rating scale questions. If you do not do this you will never know why a service is failing. Also some respondents will not feel properly consulted if there is no possibility for free expression.

Do's & don'ts in question wording

The wording of individual items is a critical aspect of a questionnaire's validity and reliability. They should aim for the simplest wording possible while still conveying the intended meaning.

- Make it short: 20 words or less per question. You can provide further clarification and context in a paragraph in smaller font below the question.
- Do not put two questions into one. Avoid questions such as "Did you like the content and appearance of the document?"
- Care should be taken to ensure that the questions are 'neutral' i.e. that they do not imply an expected or "correct" answer.
- Avoid ambiguity and negative wording. Never use double negatives (e.g. "Are you against a ban on smoking?"), which tend to confuse people.
- Technical terms and acronyms should not be used, unless you are absolutely sure that respondents know what they mean.
- If a question depends on the respondent's memory then the time period should be clearly defined and should not exceed 1 or 2 years at the most.

G. Web Statistics

Technical departments currently spend a significant proportion of their budget on information systems and web site related work.[9] GILW uses WebTrends, a commercial package, to provide web usage statistics that allow a crude estimate of the number of users at one site, which pages are the most frequently viewed or what methods the users are using to enter a particular web site, how they had arrived at the site (referring sites), etc. The system can supply some limited information on the characteristics of web site users. Domain names (and countries where they are registered) can be discovered for around 60 percent of the users of FAO sites. To access the WebTrends analysis, click on \\waicent_003\reports or email your information request to GILW.

These data are, however, difficult to interpret. A number of problems have been identified in this respect:

- Automatic indexing tools from search engines[10] can inflate the number of hits a site receives. WebTrends fortunately identifies such "robots" and it is therefore possible to deduct them from the number of users.
- Repeat use is probably underestimated as browsers will cache (store in memory) pages for re-use, so a second page view is not recorded by the website server.
- A hit does not necessarily mean that the user visiting the page has found what she/he is interested in.
- Site statistics do not include vital statistics such as gender, occupation or nationality.
- The country from which visitors are accessing the site (geographical origin of users, based on national suffixes such as .uk or .it) may overestimate the United States and to a lesser extent Europe, due to the way information requests are routed on the Internet.
- The analysis of domain suffixes (.gov, .com, etc.) typically yields little usable data since many civil servants and academics use a commercial internet access provider.

This being said, the time series of visitor numbers and duration are generally safe to use, as are the documents most downloaded.

The quality of visits may be more important than their quantity. Without some form of dialogue or feedback from users on their experience in the use of information systems, there can be no sound basis for redesign efforts.

The most powerful way to move beyond crude web statistics may be to ask users to register their e-mail addresses, gender, nationality, location, institutional affiliation and areas of interest in order to gain access to certain parts of the site providing premium content. Such a registration process, already used by some Divisions (e.g. the Climate Change and Forestry Mailing List maintained by FOP) helps identify the most interested users and get their feedback on web design and content issues, for instance through e-

[9] A recent estimate for the AG Department (Mike Robson: Evaluating Technical Department Information Systems, AGDP 2002) put the total cost at around US$10ml. per biennium, from a departmental budget of around US$90ml (core resources only).

[10] Software visiting web pages automatically in order to build up the database of pages available to search engines such as Google. Often called "spiders", "crawlers" or "robots".

mail surveys. Of course, such an approach will not reach those who did not find the site useful, a bias that should be taken into account during analysis.

H. Field Visits and Country Case Studies

Visits to a sample of countries form an important part of many external evaluations. In the context of auto-evaluation, they may be the exception rather than the rule because of their high cost. They will most probably remain confined to well-funded programme entities that have produced significant results at the country level, for instance PEs supporting pilot projects.

There are a few rules of thumb for such visits:

- they should be well planned in advance;
- there should be a small budget for local travel and hire of interpreters, etc.;
- there should be a check list of issues common to all countries, which also provides the structure for a short country report;
- the short summary report should be frank and internal to the team. Facts should be checked but the emphasis is on a comprehensive aide mémoire, not a polished report;
- the stakeholders to be seen for each case study will be identified during planning and roughly the same for each case. They may include not only those directly involved with the activities in government and as beneficiaries but also other concerned departments in government, partners in the UN system, IFIs, bilateral donors, NGOs and academic institutions, as appropriate.

The use of a common evaluation framework (checklist of issues, stakeholders, report outline) is most important, as seldom will all members of the evaluation team visit all case study sites. However, the common framework should not be regarded as static or prescriptive. The checklist of questions and group of interviewed stakeholders may evolve, provided minimum commonality is maintained.

Country visits have to make the maximum use of secondary sources, verifying through first hand visits where possible. Unfortunately there is usually very little time for site visits to field activities. Even Participatory Rural Appraisal (PRA) techniques take valuable time. Among the tools developed for PRA, the following hold useful pointers for structuring visits to project sites:

- Give guidance to the nationals organizing the site visits well beforehand on what is expected;
- Maintain a degree of flexibility in the programme to be able to follow up on new insights and not waste time on unproductive visits;
- Try and adopt a transect approach to selection of sites for visits with regard to agro-ecology, level of development and perceived success and failure; and
- In villages visited, try and ensure that the team meets not only with village leaders but also with ordinary villagers. This is facilitated by making it clear in advance that the team wishes to meet with everybody and also by the team splitting up and walking transects through the village and talking to villagers who are found in their houses or on their farms.

Commissioning case studies from local consultants is another possibility to collect specific national information, either in preparation for country visits or when country visits cannot be used. This has its limitations however. It is administratively time-consuming and requires good supervision; and in general it has been found more useful for specifically defined factual information than for analysis or insights.

I. Peer Review Panels

Peer review panels have now become a standard part of the FAO PBEE approach for evaluations submitted to the Governing Bodies whenever evaluation teams are not externally led. This assures an external input and helps to provide confirmation of the validity of the evaluation. A few rules of thumb about peer review panels:

- Panels should comprise 3 to 6 people, convened at Headquarters for about 2 to 4 days;
- Panelists should be independent, knowledgeable and respected in their field; not recently involved in FAO work; neutral on the issues at hand or at least with a balance of different views in the panel;
- Panelists are usually not paid honorarium. DSA and travel can cost from US$5,000 to 15,000 depending on the size of the panel; and
- Panels appoint their own chairperson and rapporteur.

The panel begins with a briefing on the findings of the evaluation and a question-and-answer session with the concerned managers (i.e. the people in charge of the evaluation as well as the people whose programme is being evaluated). At least a day is normally allowed for more individual meetings with concerned technical staff. The panel then prepares its own short report (2-4 pages) commenting on the findings and recommendations of the evaluation, reinforcing those points it regards as important and presenting any additional or divergent views it may have.

AE managers should be aware that peer review panels can produce a good deal of conflict when it is not made clear that they act as resource panels, not as parallel evaluation teams.

A cheaper alternative is to send the evaluation report to a wide group of implementation partners for comments. This will ensure some degree of external verification of the evaluation objectivity, but represents a much weaker process of external review. In order to emulate a real review panel, partners in the virtual review panel should share their comments with all other panelists.

ANNEXES

Annex 1: Outline for Terms of Reference and List of Evaluation Issues

Annex 2: Outline for an Auto-Evaluation Report

Annex 3: Bibliography

Annex 1: Outline for Terms of Reference

1. Background

Provides the context for the evaluation and should indicate as a minimum:

a) a description of the PE as designed: objectives, planned major outputs; starting and ending dates, budget, main inputs;

b) a brief overview of the history behind the PE;

c) a description of major activities and outputs to date; and

d) problems or emerging issues identified by management.

2. Purpose of the Evaluation

This section should briefly state why the evaluation is being held, remembering that the reason for evaluation is to provide an input to future direction. For instance:

"The evaluation is intended to provide recommendations to the department on the further steps necessary to consolidate progress and ensure achievement of objectives."

3. Scope of the Evaluation

The following represents the standard points to be included, but in any case, it is vital that the list of issues be adapted to specific concerns and questions which the concerned staff want to find answers to. Note that the list of issues is not limitative. Evaluators should always be able to raise unforeseen issues.

a) Relevance of the programme entity to development priorities and needs of Member Nations;

b) Clarity, logical consistency and realism of the programme entity design, including specification of inputs, outputs, outcomes and objectives, targets, identification of users and beneficiaries, and prospects for sustainability;

c) Realism and clarity of external institutional relationships, and in the managerial and institutional framework for implementation;

d) Efficiency of project implementation including: availability of funds and human resources as compared with budget; managerial and work efficiency; and implementation difficulties;

e) Results, including a systematic assessment of outputs produced to date (in quantity and quality) and progress towards the realization of the programme entity outcomes and objective;

f) Contributions to gender and social equity, in particular contributions to the Gender Plan of Action, and contributions to PAIAs;

g) Prospects for sustaining the results by the primary users and partners after the termination of the programme entity;

h) Cost-effectiveness of the programme entity(ies); and

i) A review of emerging issues of particular importance to management.

Add a paragraph along these lines:

> *"Based on the above analysis the evaluators will draw specific conclusions and make recommendations for any necessary further action by FAO to ensure a successful implementation of the programme entity, including opportunities that may be grasped and issues that should be resolved. The evaluation will also draw attention to any lessons of general interest."*

4. Roles in the Auto-Evaluation

This section should describe:

a) who is responsible for overseeing and coordinating the evaluation;

b) the composition and competencies of the evaluation team, including whether it is envisaged to hire external consultants; and

c) the composition and competencies of the peer review panel, if envisaged.

5. Methodology

This section should outline the main evaluation techniques to be used, including the information source (informants) and a rough time frame for the envisaged techniques. These may include, though not limited to:

a) Which indicators should be measured;

b) Desk studies and/or annotated bibliographies;

c) SWOT analysis and/or other group facilitation techniques;

d) Semi-structured interviews / focus groups

e) Questionnaire surveys;

f) Web statistics; and

g) Country case studies.

6. Evaluation Outputs

Terms of reference should describe the reporting procedure, set an indicative date for the draft report, and outline debriefing arrangements.

7. Budget

Based on the above and notably sections 4 and 5, elaborate a budget that includes staff and non-staff resources, i.e. the estimated staff time spent on managing the evaluation plus the financial resources required to secure external outputs or fund country visits.

Annex 2: Outline for an Auto-Evaluation Report

I. Executive Summary (Main Findings and Recommendations)

II. Introduction

III. Background and Context

IV. Relevance to Priorities and Needs of Member Nations

V. Assessment of Programme Entity Design

A. Clarity, consistency and realism of the programme entity design (including inputs, outputs, outcomes and objectives, users and beneficiaries, and prospects for sustainability)

B. Realism and clarity of external institutional relationships, and in the institutional framework for implementation

VI. Assessment of PE Implementation, Processes, Efficiency and Management

A. Financial and human resource management

B. Activities undertaken and outputs produced

C. Partnerships and collaborative processes

VII. Assessment of Results and Effectiveness

A. Audience of the PE and documented outcomes.

B. Progress towards the realization of the objective

C. Achievements in terms of gender and social equity

D. Cost-effectiveness

E. Major factors affecting the project results

F. Sustainability of results

VIII. Conclusions and Recommendations

IV Lessons Learned

Annexes

1. Terms of reference
2. Key persons met or interviewed
3. Documents consulted by the mission

Annex 3: Bibliography

Terminology:

The most widely accepted evaluation terminology can be found in the OECD glossary of evaluation terms, in English, French and Spanish:
http://www.oecd.org/dataoecd/29/21/2754804.pdf

Planning for evaluation:

The American Evaluators Association and the Michigan University maintain a number of checklists for evaluation management using various evaluation models. A bit too detailed for non-specialists but a useful source of ideas and tips nonetheless.
http://www.wmich.edu/evalctr/checklists/checklistmenu.htm

Indicators:

A short primer on indicator theory from UNDP: http://accsubs.unsystem.org/ccaqfb-intranet/RBB-RBM/UNDPIndicaors.pdf, as well as http://stone.undp.org/undpweb/eo/evalnet/docstore3/yellowbook/documents/key_indicators.pdf

In the 90's the World Bank invested quite a lot of efforts into its "Performance Monitoring Indicators", in effect a list of standard indicators for all development sectors (energy, micro-finance etc.). Their handbook discussing indicators is here: http://www-wds.worldbank.org/servlet/WDS_IBank_Servlet?pcont=details&eid=000009265_3961219094954

An EU paper on common questions with criteria and indicators:
http://europa.eu.int/comm/agriculture/rur/eval/evalquest/a_en.pdf

The US Environment Protection Agency has given some thoughts to the evaluation of indicators: http://www.epa.gov/indicators/grformat.htm

More bibliography on indicators:
http://www.eldis.org/participation/pme/Eldis_selection.htm

Surveys:

A Brief Guide to Questionnaire Development, by Dr. Robert Frary -
http://www.testscoring.vt.edu/fraryquest.html

A very good and short tutorial on survey design, by David S. Walonick:
http://www.statpac.com/surveys/surveys.doc or http://www.statpac.com/surveys/ (Statpac is selling a survey software as well).

A useful list of response scales (Fair / Unfair; Agree / Disagree, etc.) at http://dataguru.org/ref/survey/responseoptions.asp

A concise guide to questionnaire development, developed for surveying the usability of information technology projects: http://atwww.hhi.de/USINACTS/tutorial/quest.html

Sites on statistical analysis:
http://www.statsoft.com/textbook/stathome.html
http://obelia.jde.aca.mmu.ac.uk/resdesgn/arsham/opre330.htm

An example of a web survey in FAO (COAG Delegates survey):
http://waicent.fao.org/coagsurvey/

Qualitative approaches - semi-structured interviews and focus groups:

A good primer to qualitative evaluation techniques can be found on the Cyfer.net website of the University of Arizona (apparently being restructured), in particular "Using Focus Groups for Evaluation", by Mary Marczak & Meg Sewell, and "The Use of Qualitative Interviews in Evaluation", by Meg Sewell.

Evaluating programme outcomes:

The IDRC evaluation office has developed a detailed methodology called "outcome mapping" to help programme evaluators and staff review their effectiveness.
http://web.idrc.ca/ev.php?url_id=26586&url_do=do_topic&url_section=201&reload=1060162955

The UNDP Evaluation Handbook, largely relying on outcome mapping:
http://stone.undp.org/undpweb/eo/evalnet/docstore3/yellowbook/

Evaluation of websites:

GILW WebTrends analysis: \\waicent_003\reports.

Hope N. Tillman: Evaluating Quality on the Net. Babson College, Massachusetts: http://www.hopetillman.com/findqual.html#my

Smith, Alastair G. "Testing the Surf: Criteria for Evaluating Internet Information Resources." The Public-Access Computer Systems Review 8, no. 3 (1997):
http://info.lib.uh.edu/pr/v8/n3/smit8n3.html

Peer Reviews

OECD-DAC has a rich collection of evaluation guidelines, including: Peer Review: a Tool for Co-Operation and Change - an Analysis of an OECD Working Method by Fabrizio Pagani: http://www.oecd.org/dataoecd/33/16/1955285.pdf

Same document in French: L'Examen par les Pairs : un Instrument de Coopération et de Changement. http://www.oecd.org/dataoecd/33/18/1955301.pdf

Food and Agriculture Organization of the United Nations
Evaluation Service (PBEE)
Viale delle Terme di Caracalla
00100 Rome, Italy
http://www.fao.org/pbe/pbee/
Comments should be sent to: evaluation@fao.org