

# **Review and Analysis of Training Impact Evaluation Methods, and Proposed Measures to Support a United Nations System Fellowships Evaluation Framework**

Prepared for the WHO's Department of Human Resources for Health  
on behalf of the UN Task Force on Impact Assessment of Fellowships

by **Michael A. Zinovieff with the support of Arie Rotem**

Geneva, July 2008

Review and analysis of training impact evaluation methods, and proposed  
measures to support a UN system fellowships evaluation framework

**Table of contents**

<b>I. INTRODUCTION.....</b>	<b>1</b>
<b>II. WHAT IS EVALUATION?.....</b>	<b>2</b>
1. DEFINITION .....	2
2. DIMENSIONS.....	4
3. GOALS.....	4
4. APPROACHES .....	5
5. PURPOSES .....	5
6. TYPES.....	6
<i>i. Formative.....</i>	<i>7</i>
<i>ii. Summative.....</i>	<i>7</i>
<i>iii. Confirmative.....</i>	<i>8</i>
<i>iv. Meta.....</i>	<i>8</i>
<i>v. Goal-based.....</i>	<i>9</i>
<i>vi. Process-based.....</i>	<i>9</i>
<i>vii. Outcomes-based.....</i>	<i>9</i>
7. PROGRAMME EVALUATION .....	10
8. MODELS/METHODS .....	11
<i>i. Objective/goal - based.....</i>	<i>12</i>
a) Donald Kirkpatrick's 4 levels.....	12
b) Jack Phillips Return on Investment (ROI).....	14
c) Hamblin's 5 levels.....	15
d) Guskey's critical levels.....	15
e) Indiana University taxonomy.....	16
f) Industrial Society stages.....	16
g) Kearns and Miller KPMT model.....	17
h) Nine outcomes model.....	17
i) Organizational elements model.....	18
j) Contemporary ROI models.....	18
<i>ii. Systems-based.....</i>	<i>19</i>
a) CIPP (Context, Input, Process, Product).....	19
b) IPO (Input, Process, Output).....	20
c) TVS (Training Valuation System).....	20
d) Pulley's responsive evaluation model.....	20
e) E-Learning models.....	21
f) Duignan's framework for outcomes systems.....	21
g) United Way of America outcomes evaluation.....	22
h) Charities Evaluation Service outcomes triangle.....	22
<i>iii. Additional evaluation methods, including Contribution Analysis.....</i>	<i>22</i>

9. CRITIQUE OF THE KIRKPATRICK MODEL .....	23
10. FEASIBILITY OF TRAINING IMPACT ASSESSMENT/SOME SUCCESS STORIES.....	25
<b>III. MEASURES TO SUPPORT AN EVALUATION FRAMEWORK .....</b>	<b>27</b>
1. CURRENT STATE OF FELLOWSHIPS AND CHANGING EXPECTATIONS .....	27
2. FELLOWSHIPS EVALUATION IN SELECTED BILATERAL AND MULTILATERAL ORGANIZATIONS .....	28
3. THE OBJECTIVES-BASED WORLD BANK (INDEPENDENT EVALUATION GROUP) APPROACH .....	30
4. THE FIVE REQUIREMENTS FOR FELLOWSHIPS EVALUATION .....	31
5. KEY FINDINGS.....	33
<b>IV. CONCLUSIONS AND RECOMMENDATIONS .....</b>	<b>35</b>
<b>V. ANNEX ON UNEG’S DOCUMENT ON EVALUATION IN THE UN SYSTEM.....</b>	<b>37</b>

## I. Introduction

Evaluation is hardly a new subject for discussion in the UN system and it is increasingly becoming a critical one. The UNDP, for instance, recently reviewed the topic at depth during regional workshops in 2006-07. Among the conclusions drawn was that:

“The enabling conditions for evaluation across UNDP remain weak. These include the variable quality of frameworks for monitoring and evaluation, the lack of systematic results monitoring, the strategic allocation of human and financial resources, and the mechanisms for quality assurance. Evaluation itself continues to be poorly organized and funded across the organization.”

It was also noted that evaluation was not adequately carried out throughout the UN system. A recommendation was made accordingly:

“All efforts should be made to ensure that the monitoring support and evaluation function is harmonized with other UN agencies, and aligned with national systems. UNDP should promote joint evaluations with UN agencies in particular.”<sup>1</sup>

In a similar vein the 16<sup>th</sup> Meeting of Senior Fellowship Officers (Paris, 6-8 November 2006) recommended the creation of a Task Force on Impact Assessment of Fellowships. Following a meeting with consultants at WHO in Geneva, 28-30 April 2008, the specific objectives of the Task Force were identified as follows:

- 1) To undertake a literature search, document analysis and a critical review of methods and processes of training impact evaluation, with a view to determine the relevance of these approaches to the assessment of the impact of UN Fellowships programmes;
- 2) With reference to the review of the literature and in consultation with the Task Force, to draft a generic evaluation framework that defines the scope, dimensions and core indicators for evaluating the impact of UN Fellowships programmes;
- 3) To identify necessary conditions and supportive measures to enable implementation of the impact evaluation framework in the context of the UN Fellowships programmes and to present the findings of this review for discussion and review at the 17<sup>th</sup> Meeting of SFOs.

The present report, derived from a variety of sources on evaluation, takes prime responsibility for the review and analysis of the literature on the subject (1<sup>st</sup> objective) and will also attempt to identify organizational “measures” which could support and enhance an evaluation framework for UN agency fellowships. Its sources are identified through a considerable number of footnotes. The report makes little claim to original authorship and should be viewed as a compilation and synopsis of the work of evaluation specialists.

A fairly obvious note of caution: evaluation refers to a wide range of activities, processes, products, etc. An astounding volume of literature is devoted to training evaluation but not much of it concerns “fellowships”. Our fellows are not trainees in the typical organizational sense and they leave our environments after their fellowships/training so we are compelled to try to measure their reaction, learning, behaviour, change, etc. from afar. Hence our evaluatory task is made infinitely more difficult than is true for, say, the

---

<sup>1</sup> See [www.undp.org/eo/documents/workshop/regional\\_ws/wsr-summary.pdf](http://www.undp.org/eo/documents/workshop/regional_ws/wsr-summary.pdf).

typical corporation which trains its employees and can measure the training results on the spot as it were.

## II. What is evaluation?

### 1. Definition<sup>2</sup>

Evaluation of training and fellowships in the UN system has apparently not been carried out at a level that will adequately measure the impact of training/fellowships or other performance improvement interventions, at least if one is to judge from the frequently expressed frustration by the UN agencies' senior management and by the agencies' constituents. Yet, systematic evaluation can provide the information needed for continuous improvement. Moreover, today managers are no longer satisfied with knowing how many fellows underwent training, how they liked it, and what they learned. Increasingly managers want to know if the fellows are using what they learned, and – most importantly – what if any institutional results were improved.

In any review of evaluation it is first essential to define the term itself, as well as its stakeholders and its goals. Then an analysis can be made of the various types of evaluation and major models/methodologies commonly applied to measure impact.

Probably the most frequently given definition is:

*Evaluation is the systematic assessment of the worth or merit of some object*<sup>3</sup>

The definition is hardly perfect. There are many types of evaluations that do not *necessarily* result in an assessment of worth or merit – descriptive studies, implementation analyses, and formative evaluations, to name a few. Better perhaps is a definition that emphasizes the information-processing and feedback functions of evaluation. For instance, one might say:

*Evaluation is the systematic acquisition and assessment of information to provide useful feedback about some object*<sup>4</sup>

For the American Evaluation Association evaluation involves assessing the strengths and weaknesses of programmes, policies, personnel, products and organizations to improve their effectiveness.

---

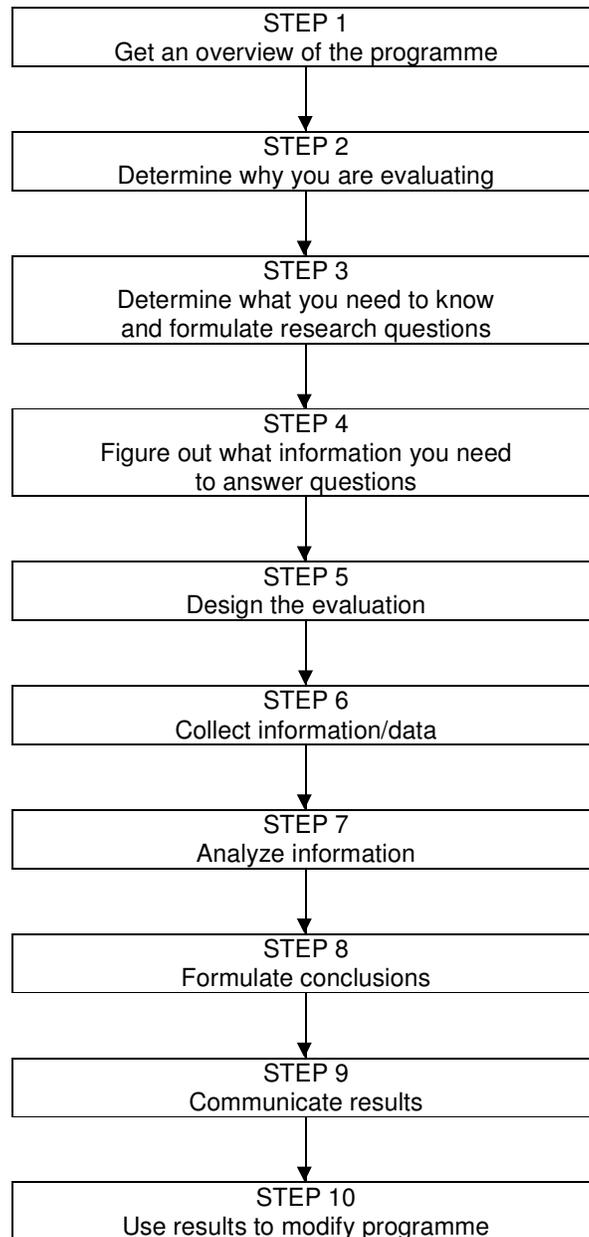
<sup>2</sup> Evaluation literature has a jargon of its own but even among the specialists terminology can differ. What is “purpose” for one may be “type” for another. To provide a perspective on evaluation, somewhat different than the present synopsis, Paul Duignan’s “Introduction to Strategic Evaluation” at <http://www.strategicvaluation.info/documents/104.htm> is recommended reading (8 pages) for its definition of evaluation approaches, purposes, methods and designs.

For a very useful on line “course” on the general subject see also “Introduction to Evaluation”, the R561 course on evaluation and change by the Instructional Systems Technology Dept. of Indiana University’s School of Education at [www.indiana.edu/~istr561/knuth06sum/unit1a.shtml](http://www.indiana.edu/~istr561/knuth06sum/unit1a.shtml).

<sup>3</sup> See [www.socialresearchmethods.net/kb/intreval.htm](http://www.socialresearchmethods.net/kb/intreval.htm).

<sup>4</sup> See William Trochim, currently President of the American Evaluation Association, at [www.eval.org/part\\_1.asp](http://www.eval.org/part_1.asp).

Another definition sees evaluation as the systematic collection and analysis of data needed to make decisions (see Lana Muraskin, “Understanding Evaluation, the Way to Better Prevention Programs”). With all the slight differences in each definition, several distinct “steps” are usually followed in any evaluation.<sup>5</sup>



It should be noted that most definitions emphasize acquiring and assessing information rather than assessing worth or merit because all evaluation work involves collecting and sifting through data, making judgements about the validity of the information and of inferences we derive from it, whether or not an assessment of worth or merit results.

---

<sup>5</sup> See [www.evaluationwiki.org/wiki/index.php/Evaluation\\_Definition](http://www.evaluationwiki.org/wiki/index.php/Evaluation_Definition).

This feedback is provided to the stakeholders, i.e. anyone without whose input an organization would be unable to function. In evaluation the term stakeholder can be more broadly defined as one who may have no formal role in an organization's functions but still be positively or negatively affected by its functioning. The goals of evaluation are to influence decision-making or policy formulation through the provision of empirically-driven feedback. It is imperative to ensure that all stakeholders are included in the evaluation process before a particular evaluation type or model/methodology is chosen.

Indeed, one evaluation model (or method as some prefer) focuses essentially on the stakeholders, arguing that to adequately evaluate training it is necessary to assess the extent to which all stakeholder groups are satisfied with what they have given to and received from the training.<sup>6</sup>

## 2. Dimensions

Evaluation literature refers to the “dimensions of evaluation” as process, outcome and impact. These concepts are fundamental and we will return to them in other contexts more fully.

- *Process evaluations*  
Process Evaluations describe and assess programme materials and activities. Establishing the extent and nature of programme implementation is an important first step in studying programme outcomes; that is, it describes the interventions to which any findings about outcomes may be attributed. Outcome evaluation assesses programme achievements and effects.
- *Outcome evaluations* (see also 8.ii.f and g)  
Outcome Evaluations study the immediate or direct effects of the programme on participants. The scope of an outcome evaluation can extend beyond knowledge or attitudes, however, to examine the immediate behavioural effects of programmes.
- *Impact evaluations*  
Impact Evaluations look beyond the immediate results of policies, instruction, or services to identify longer-term as well as unintended programme effects. Very useful reports on this subject have notably been made by the Center for Global Development<sup>7</sup>, and by Deloitte Insight Economics.<sup>8</sup> For a comprehensive review of the three dimensions – process, outcome and impact – see [www.evaluationwiki.org/wiki/index.php/Evaluation\\_Definition](http://www.evaluationwiki.org/wiki/index.php/Evaluation_Definition), noted above.

## 3. Goals<sup>9</sup>

The generic goal of most evaluations is thus to provide useful feedback to a variety of audiences including sponsors, donors, client-groups, administrators, staff, and other relevant constituencies. Most often, feedback is perceived as “useful” if it aids in decision-making. But the relationship between an evaluation and its impact is not a simple one – studies that seem critical sometimes fail to influence short-term decisions, and studies that

---

<sup>6</sup> See Fred Nichols “A Stakeholder Approach to Evaluation”.

<sup>7</sup> See “When Will We Ever Learn: Improving Lives through Impact Evaluation”, May 2006 publication by the Center for Global Development.

<sup>8</sup> See “Impact Monitoring and Evaluation Framework”, June 2007 report at [www.crcwood.unimelb.edu.au/docs/CRCA\\_Framework\\_Background.pdf](http://www.crcwood.unimelb.edu.au/docs/CRCA_Framework_Background.pdf).

<sup>9</sup> See inter alia [www.socialresearchmethods.net/kb/intreval.htm](http://www.socialresearchmethods.net/kb/intreval.htm), above.

initially seem to have no influence can have a delayed impact when more congenial conditions arise. Despite this, there is broad consensus that the major goal of evaluation should be to influence decision-making or policy formulation through the provision of empirically-driven feedback.

#### 4. Approaches

An evaluation approach is a general way of looking at or conceptualizing evaluation; the main evaluation approaches according to Paul Duignan (“Introduction to Strategic Evaluation”) include notably:

- Utilisation-focused evaluation – determines methods on the basis of what is going to be most useful to different audiences;
- Empowerment evaluation – emphasises that the evaluation process and methods should be empowering to those who are being evaluated;
- Stakeholder evaluation – looks at the differing perspectives of all of a programme’s stakeholders (those who have an interest in it);
- Goal-free evaluation – in which the evaluator’s task is to examine all of the outcomes of a programme, not just its formal outcomes as identified in its objectives;
- Naturalistic or 4<sup>th</sup> generation evaluation – emphasises the qualitative uniqueness of programmes and is a reaction against the limitation of quantitative evaluation approaches;
- Theory based evaluation – puts an emphasis on detailing the assumptions on which a programme is based (intervention logic) and follows those steps to see if they occur;
- Strategic evaluation – emphasises that evaluation design decisions should be driven by the strategic value of the information they will provide for solving social problems.

#### 5. Purposes

There are various ways of describing various *purposes* of evaluation activity, e.g. design, developmental, formative, implementation, process, impact, outcome and summative. The evaluation purpose is best understood as identifying what evaluation activity is going to be used for. Recent years have seen evaluation move to develop types of evaluation that are of use right across a programme lifecycle. It should be noted that any particular evaluation activity can have more than one purpose.

The range of evaluation terms are used in various ways in the evaluation literature. A common way of defining them is as follows (see Duignan):

- Design, developmental, formative, implementation – evaluative activity designed to improve the design, development, formation and implementation of a programme;
- Process – evaluation to describe the process of a programme. Because the term process could conceivably cover all of a programme from its inception to its outcomes, it is conceptually useful to limit the term process evaluation to activity describing the programme during the course of the programme, i.e. once it has been initially implemented;

- Impact, outcome and summative – looking at the impact and outcome of a programme, and in the case of summative, making an overall evaluative judgment about the worth of a programme.

The purposes of evaluation also relate to the intent of the evaluation:<sup>10</sup>

- Gain insight – provide the necessary insight to clarify how programme activities should be designed to bring about expected changes;
- Change practice – improve the quality, effectiveness, or efficiency of programme activities;
- Assess effects – examine the relationship between programme activities and observed consequences;
- Affect participants – use the processes of evaluation to affect those who participate in the inquiry. The systematic reflection required of stakeholders who participate in an evaluation can be a catalyst for self-directed change. Evaluation procedures themselves will generate a positive influence.

Some organizations work towards the Investors in People (IIP) initiative which defines the purposes of evaluation as follows:<sup>11</sup>

- IIP Indicator 4.1      The organization evaluates the impact of training and development on knowledge, skills and attitude
- IIP Indicator 4.2      The organization evaluates the impact of training and development actions on performance
- IIP Indicator 4.3      The organization evaluates the contribution of training and development to the achievement of its goals and targets

In the general discussion on the purposes of evaluation a note of caution is often introduced with respect to cost vs. benefit. This is known as “rationalizing evaluation”. It is important not to get “carried away” with an evaluation effort which is disproportionately greater than the investment made, or the benefit likely to be achieved. The IPD (Independent Professional Development, London) study “Making Training Pay” (1997) suggests that the scope of an evaluation strategy should be carefully weighed against the following considerations:

- The size of the training investment
- The number of staff involved
- The likelihood that the training will be repeated
- The criticality of the training to the business
- The “newness” of the training methods used.

## 6. Types<sup>12</sup>

The most basic difference is between what is known as the formative and the summative types of evaluation. In more recent years the concepts of confirmative and meta evaluation have received much attention as well.

<sup>10</sup> See “University of British Columbia e-learning course”.

<sup>11</sup> See inter alia [www.dba.co.uk/tips/vol3/vol3iss5.htm](http://www.dba.co.uk/tips/vol3/vol3iss5.htm).

<sup>12</sup> See also [www.evaluators-webring.net/Independent\\_evaluators\\_webring\\_definitions\\_May06.pdf](http://www.evaluators-webring.net/Independent_evaluators_webring_definitions_May06.pdf).

Formative evaluations strengthen or improve the object being evaluated – they help form it by examining the delivery of the programme or technology, the quality of its implementation, and the assessment of the organizational context, personnel, procedures, inputs, and so on. Summative evaluations, in contrast, examine the effects or outcomes of some object – they summarize it by describing what happens subsequent to delivery of the programme or technology; assessing whether the object can be said to have caused the outcome; determining the overall impact of the causal factor beyond only the immediate target outcomes; and, estimating the relative costs associated with the object.

i. Formative

Formative evaluation includes several evaluation types:

- *Needs assessment* determines who needs the programme, how great the need is, and what might work to meet the need;
- *Evaluability assessment* determines whether an evaluation is feasible and how stakeholders can help shape its usefulness;
- *Structured conceptualization* helps stakeholders define the programme or technology, the target population, and the possible outcomes;
- *Implementation evaluation* monitors the fidelity of the programme or technology delivery;
- *Process evaluation* investigates the process of delivering the programme or technology, including alternative delivery procedures.

ii. Summative

Summative evaluation can also be subdivided:

- *Outcome evaluations* investigate whether the programme or technology caused demonstrable effects on specifically defined target outcomes;
- *Impact evaluation* is broader and assesses the overall or net effects – intended or unintended – of the programme or technology as a whole;
- *Cost-effectiveness and cost-benefit analysis* address questions of efficiency by standardizing outcomes in terms of their dollar costs and values;
- *Secondary analysis* re-examines existing data to address new questions or use methods not previously employed;
- *Meta-analysis* integrates the outcome estimates from multiple studies to arrive at an overall or summary judgement on an evaluation question.

There is an abundance of literature on the subject of formative and summative evaluations; for a succinct presentation, which also identifies the questions and methodologies addressed under these types, see [www.socialresearchmethods.net/kb/intreval.htm](http://www.socialresearchmethods.net/kb/intreval.htm).

To a number of evaluation specialists these two fundamental types of evaluation do not suffice to constitute what is referred to as “full-scope evaluation”.<sup>13</sup>

Full-scope evaluation systematically judges the merit and worth of a long-term training programme before, during, and after implementation. Full-scope

---

<sup>13</sup> See notably Joan Dessinger and James Morley, “Full-Scope Evaluation: Raising the Bar”, 2003.

evaluation is appropriate only for training programmes that are designed to run for one year or more; it is not appropriate for a one-time training event, such as a single-session workshop to introduce a new product to sales representatives.

Full-scope evaluation integrates four types of programme evaluation – formative, summative, confirmative and meta – into the training programme evaluation plan. Working together, the four types of evaluation help to determine the value of a long-term training programme and develop the business case or rationale for maintaining, changing, discarding, or replacing the programme. Full-scope evaluation introduces the concepts of the confirmative and the meta type of evaluation. These concepts may not be of priority interest to other than evaluation practitioners but deserve at least the brief mention below.

iii. Confirmative

Confirmative evaluation goes beyond formative and summative evaluation; it moves traditional evaluation a step closer to full-scope evaluation. During confirmative evaluation, the evaluation and training practitioner collects, analyzes, and interprets data related to behaviour, accomplishment, and results in order to determine the continuing competence of learners or the continuing effectiveness of instructional materials and to verify the continuous quality improvement of education and training programmes.

While formative and summative evaluations comprise two initial levels, confirmative evaluation assesses the transfer of learning to the “real world”:

- a) *Level one*: evaluate programmes while they are still in draft form, focusing on the needs of the learners and the developers;
- b) *Level two*: continue to monitor programmes after they are fully implemented, focusing on the needs of the learners and the programme objectives;
- c) *Level three*: assess the transfer of learning to the real world.

Even level four of Kirkpatrick’s four levels of evaluation is confirmative evaluation by another name. Level four measures the results of training in terms of change in participant behaviour and tangible results that more than pay for the cost of training.<sup>14</sup>

iv. Meta<sup>15</sup>

Formative, summative, and confirmative evaluation are all fodder for meta evaluation. Meta evaluation is all about evaluating the evaluation. The evaluator literally zooms in on how the evaluation was conducted. The purpose of meta evaluation is to validate the evaluation inputs, process, outputs, and outcomes. It serves as a learning process for the evaluator and makes the evaluators accountable.

There are two types of meta evaluation: type one and type two. Type one meta evaluation is conducted concurrently with the evaluation process. It is literally a formative evaluation of evaluation. Type two meta evaluation is the more

---

<sup>14</sup> See Dessinger and Morely “Confirmative Evaluation”, 2003.

<sup>15</sup> See inter alia Peter Wiles “Meta-evaluation”, 2004.

common approach. It is conducted after formative, summative, and at least one cycle of confirmative evaluation is completed. Some evaluation specialists have also defined the several types of evaluation more thematically, as below.

v. Goal-based<sup>16</sup>

Goal-based evaluations are evaluating the extent to which programmes are meeting predetermined goals or objectives. Questions to ask when designing an evaluation to see if the goals have been reached include:

- How were the programme goals (and objectives, if applicable) established?
- Was the process effective?
- What is the status of the programme's progress toward achieving the goals?
- Will the goals be achieved according to the timelines specified in the programme implementation or operations plan? If not, then why?
- Do personnel have adequate resources (money, equipment, facilities, training, etc.) to achieve the goals?

vi. Process-based<sup>17</sup>

Process-based evaluations are geared to fully understanding how a programme works – how does it produce that results that it does. These evaluations are useful if programmes are long-standing and have changed over the years, employees or customers report a large number of complaints about the programme, there appear to be large inefficiencies in delivering programme services and they are also useful for accurately portraying to outside parties how a programme operates.

There are numerous questions that might be addressed in a process evaluation. These questions can be selected by carefully considering what is important to know about the programme. Examples of questions include:

- On what basis do employees and/or the customers decide that products or services are needed?
- What is required of employees in order to deliver the product or services?
- How are employees trained about how to deliver the product or services?
- How do customers or clients come into the programme?
- What is required of customers or client?

vii. Outcomes-based<sup>18</sup>

Evaluation with an outcomes focus is increasingly important for nonprofits and asked for by funders. An outcomes-based evaluation tries to ascertain if the organization is really doing the right programme activities to bring about the outcomes it believes to be needed by its clients. Outcomes are benefits to clients from participation in the programme. Outcomes are usually in terms of enhanced learning (knowledge, perceptions/attitudes or skills) or conditions, e.g. increased literacy, self-reliance, etc. Outcomes are often confused with programme outputs or units of services, e.g. the number of clients who went through a programme.

The United Way of America ([www.unitedway.org/outcomes/](http://www.unitedway.org/outcomes/)) provides an excellent overview of outcomes-based evaluation, including introduction to

---

<sup>16</sup> See inter alia [http://www.managementhelp.org/evaluatn/fnl\\_eval.htm](http://www.managementhelp.org/evaluatn/fnl_eval.htm).

<sup>17</sup> See inter alia [http://www.managementhelp.org/evaluatn/fnl\\_eval.htm](http://www.managementhelp.org/evaluatn/fnl_eval.htm).

<sup>18</sup> See inter alia [http://www.managementhelp.org/evaluatn/fnl\\_eval.htm](http://www.managementhelp.org/evaluatn/fnl_eval.htm).

outcomes measurement, a programme outcome model, why to measure outcomes, use of programme outcome findings by agencies, eight steps to success for measuring outcomes, examples of outcomes and outcome indicators for various programmes and the resources needed for measuring outcomes. (See also section 8.ii.g below).

## 7. Programme evaluation

Before reviewing the main thematic “models” of evaluation, a few words on the interlinking of “goals”, “objectives” and “programme”. In evaluation literature “programme evaluation” is often used. It is more than an evaluation “type” but neither is it a “model” per se. Michael Patton in particular has written extensively on the subject<sup>19</sup>. Broadly, programme evaluation is a comprehensive form of ascertaining to what extent goals/objectives have been achieved. To effectively conduct programme evaluation one first needs to have “programme” (a strong impression of what customers/clients actually want and need).

Programme evaluation is carefully collecting information about a programme or some aspect of a programme to make necessary decisions about it. Programme evaluation can include any or a variety of at least 35 different types of evaluation (according to Patton, others have identified even more). The type and model of evaluation one undertakes to improve one’s programmes depends on what one wants to learn about the programme. One should worry less about what type/model of evaluation one needs and worry more about what one needs to know to make the programme decisions one needs to make, and worry also about how one can accurately collect and understand that information.

Patton notes that among the key questions to consider when designing a programme evaluation the following are of priority:

- 1) For what purposes is the evaluation being done, i.e. what do you want to be able to decide as a result of the evaluation?
- 2) Who are the audiences for the information from the evaluation, e.g. customers, bankers, funders, board, management, staff, customers, clients, etc?
- 3) What kinds of information are needed to make the decision you need to make and/or enlighten your intended audiences, e.g. information to really understand the process of the product or programme (its inputs, activities and outputs), the customers or clients who experience the product or programme, strengths and weaknesses of the product or programme, benefits to customers or clients (outcomes), how the product or programme failed and why, etc.
- 4) From what sources should the information be collected, e.g. employees, customers, clients, groups of customers, or clients and employees together, etc.
- 5) How can that information be collected in a reasonable fashion, e.g. questionnaires, interviews, examining documentation, observing customers or employees, conducting focus groups among customers or employees, etc.
- 6) When is the information needed (so, by when must it be collected)?
- 7) What resources are available to collect the information?

---

<sup>19</sup> See inter alia his “Utilization-Focused Evaluation”, 1997.

The writings of Patton stress the need to focus on “programmes” (goals/objectives) before initiating an evaluation. To undertake evaluation one must also choose an appropriate “type”, as identified above, and then determine if anyone “model” or “method”, and/or combination of such, best fits one’s situation. We will therefore now describe some major “models” thematically.

## 8. Models/methods

The thematic categorization of evaluation types is echoed in the literature with respect to the various evaluation “models” (some prefer the term “methods”) which is of prime importance in the effort to identify an evaluation framework. Discussions concerning the definition, dimensions, goals, approaches, purposes, and types of evaluation may be helpful but the crux of an evaluation “framework” (Objectives 2 and 3 of the TOR of the Task Force) lies in determining, as far as it is feasible, the respective relevance of these models to the assessment of the impact of UN system fellowships. There are a myriad of such models for which reason a thematic classification of these would undoubtedly facilitate the review. A recent article on “approaches to evaluation” by Deniz Eseryel identifies six such general approaches:

- Goal-based evaluation
- Goal-free evaluation
- Responsive evaluation
- Systems evaluation
- Professional review
- Quasi-legal

The Indiana University website above defines these approaches as follows:

- Goal-based evaluation begins with goals in mind and seeks to determine if those goals were achieved;
- Goal-free evaluation does not seek to confirm or deny a pre-determined outcome or goal. Rather, it seeks to discover any benefits that result from the intervention;
- Responsive evaluation is an approach that it is based on client requirements. This can present unique challenges for the evaluator, but it is a common approach;
- The systems approach to evaluation focuses on whether the intervention was efficient and effective;
- Professional review evaluation uses external expert appraisal to evaluate instead of other commonly used and accepted methods;
- The quasi-legal approach is infrequently practiced, but it uses an actual court-of-inquiry format to present evidence, take testimonials, and evaluate an intervention or product.

Generally, however, the literature focuses essentially on goal or objective-based vs. systems-based models. Goal-based models (such as the “bible” of evaluation models, Donald Kirkpatrick’s “Evaluation Training Programs”) may help.

Practitioners think about the purposes of evaluation ranging from purely technical to covertly political purpose. However, these models do not define the steps necessary to achieve purposes and do not address the ways to utilize results to improve training. The difficulty for practitioners following such models is in selecting and implementing appropriate evaluation methods (quantitative, qualitative, or mixed). Naturally, many

organizations do not use the entire model, and training ends up being evaluated only at the reaction, or at best, at the learning level. As the level of evaluation goes up, the complexities involved increase. This may explain why only levels 1 and 2 are used. More on the Kirkpatrick's model below.

On the other hand, system-based models (e.g. CIPP, IPO and TVS) seem to be more useful in terms of thinking about the overall context and situation but they may not provide sufficient granularity. Systems-based models may not represent the dynamic interactions between the design and the evaluation of training. Few of these models provide detailed descriptions of the processes involved in each steps. None provide tools for evaluation. Furthermore, these models do not address the collaborative process of evaluation, that is, the different roles and responsibilities that people may play during an evaluation process. More on these models below.

i. Objective/goal - based

a) Donald Kirkpatrick's 4 levels<sup>20</sup>

Discussion on the subject of evaluation types may appear somewhat academic. However, in evaluation literature this discussion inevitably leads to the very concrete examples of evaluation models and schemes. The most famous – and applied – evaluation model was developed by Donald J. Kirkpatrick (notably in his “Evaluating Training Programs”). Kirkpatrick described 4 levels of training evaluation: reaction, learning, behaviour and results. He identified the four levels as:

- *Reaction* – a measure of satisfaction (what the trainees/fellows thought and felt about the training); evaluation here focuses on the reaction of individuals to the training or other improvement intervention;
- *Learning* – a measure of learning (the resulting increase in knowledge or capability); evaluation here assesses what has been learned as measured with end of course tests;
- *Behaviour* – a measure of behaviour change (extent of behaviour and capability improvement and implementation/application); evaluation here measures the transfer of what has been learned back to the workplace;
- *Results* – a measure of results (the effects on the institutional environment resulting from the fellows' performance); evaluation here measures (at least tries to) the impact of the training on overall organizational results (in the private sector on business results).

In the framework of the above summary of “types” of evaluation levels 1 and 2 are normally seen as part of formative evaluation, whereas levels 3 and 4 are typically associated with summative evaluation. There have also been attempts to establish a level 5 by measuring the impact at a societal level (in business terms, by calculating return on investment (ROI). Levels 4 and 5 are associated with normative and/or meta evaluation to achieve an ideal “full-scale” evaluation.

---

<sup>20</sup> See notably 1998 edition of his book “Evaluating Training Programs: The Four Levels”.

Although most organizations have some form of level 1 evaluations for their training programmes, the number diminishes as the levels increase. And very few organizations take advantage of the rewards of level 4 evaluations. The following should be borne in mind:

Level 1 (reaction) and level 2 (knowledge and skills) evaluations can lead to a false sense of security; there may be no relationship between how participants feel about the training and improved individual and organizational performance; level 3 evaluations can be used to refine the training provided, but level 4 will determine whether it has value. It may not be desirable, practical, or necessary to do all levels of evaluation. Each organization needs to select the level that will produce the information required to evaluate the fellowship programme.

Again, Kirkpatrick’s model consists of 4 levels that progress in difficulty from 1 (the easiest to conduct) to 4 (the hardest). When choosing the appropriate model to include in an organizational assessment, it is essential first to identify the questions the evaluation needs to address. Kirkpatrick expressed this in a tabulation<sup>21</sup>:

Level	Measurement focus	Questions addressed
1 - Reaction	Trainees's perceptions	What did trainees think of this training?
2 - Learning	Knowledge/skills gained	Was there an increase in knowledge or skill level?
3 - Behaviour	Worksite implementation	Is new knowledge/skill being used on the job?
4 - Results	Impact on organization	What effect did the training have on the organization?

Level 1 (Reaction) is the most commonly-used method of evaluation, probably because it is the easiest to administer and evaluate. This level produces what has been dubbed the “smile sheet”, which measures how well the trainees like the training. Level 2 (Learning) is not as well-used in business settings as an evaluation technique; public sector/academic settings are more likely to use level 2 techniques; these are most reliable when pre- and post- evaluations are utilized. Few UN agencies have gone beyond the first two levels. And in this (understandable, no doubt) failure lies much of the current sense of dissatisfaction with evaluation exercises. The fact is that in both private and public sectors there is today an increasing need to show concrete evidence that training/fellowships are achieving their goals of changing behaviour on the job (level 3) and are also contributing to the institutional “bottom” line. The problem is that trainers will probably not do levels 3 and 4 evaluations unless they are told to do so. Level 3 evaluations are difficult because human behaviour needs to be measured. Some believe level 4 evaluations may actually be easier to accomplish than level 3, since level 4 is (at least ideally) tied to measurable information. Some trainers therefore believe that a positive level 3 evaluation implies success at level 4. For UN agencies, the issue is whether they are prepared to spend the money to carry out genuine evaluation exercises

<sup>21</sup> For an excellent synopsis of Kirkpatrick’s model, inclusive of comparative “grids” and tabulations, see also [www.businessballs.com/kirkpatricklearningevaluationmodel.htm](http://www.businessballs.com/kirkpatricklearningevaluationmodel.htm).

which go beyond the first two levels. If it is, they would then need to decide on the most appropriate evaluation model/methodology.

In fact, research from the American Society of Training and Development<sup>22</sup> shows that over 75% of organizations measure only the level of reaction through the use of questionnaires, the “smile” or “happy sheets”.

The ASTD study also showed (2002) that 44% of US corporations measured how much people learned, while only 21% how much people changed their behaviour as a result of training; indeed a mere 11% measured whether training affected organizational results. This is a danger for the quality of training, of course, for if participant reaction is the only measure of performance too much energy, on the part of trainers, could be devoted to obtaining favourable ratings and not helping people to learn, grow, and change (for the better). In this context critics of the Kirkpatrick model (more on this later) note that just as there is no proven causal link between reaction and learning, there is no such link between learning and behaviour change. Just because a participant learned something does not mean anything will be done with the learning – hence the importance of behaviour change evaluation well after the training and, to some, recourse to different evaluation models/methods.<sup>23</sup>

b) Jack Phillips Return on Investment (ROI)<sup>24</sup>

In his many books and articles Phillips has gone beyond even Kirkpatrick’s level 4 to focus on real measurement of ROI (justification of the cost of training based on the return on investment and organizational impact). Training in this sense has thus moved from satisfying trainees to improving organizational performance. Training/fellowships are carried out to have a positive impact on the organization. This is obviously a far cry from the “smile sheets” forming the basis of level 1 evaluation. Today, many evaluators point out that while the Kirkpatrick’s model is useful to evaluate a) whether learners liked their instruction, b) whether they learned something from it, and c) whether it had some positive effect for the organization, its weakness is that it cannot be used to determine the cost-benefit ratio of training (ROI). These modern evaluators have consequently recommended adding the so-called fifth level to Kirkpatrick’s model, at least for some programmes. This may be too much for a UN agency since it requires collecting level 4 data, converting the results to monetary values, and then comparing those results with the cost of the training/fellowship programme. There is even a basic formula for calculating ROI, not identified here for its apparent incompatibility with the constraints inherent for UN agencies.

Research into the training evaluation models that have been proposed over the last 45 years since Kirkpatrick’s framework, show that many have used the four levels as a basis for their thinking, though Phillips has also had “impact”.

---

<sup>22</sup> See the ASTD “Training & Development Handbook”, edited by Robert Craig, which inter alia contains a chapter by Kirkpatrick on “Evaluation” and a chapter by Jack Phillips on “Measuring the Results of Training” but beware – the Handbook runs well over one thousand pages.

<sup>23</sup> See notably Alliger & Janak, “Kirkpatrick’s Level of Training Criteria: 30 years later”, 1989.

<sup>24</sup> See notably 2003 edition of his “Return on Investment in Training and Performance Improvement Programs”.

A few of these (goal-based) models/methods are identified below:<sup>25</sup>

c) Hamblin's 5 levels<sup>26</sup>

Hamblin was one of the first to modify Kirkpatrick's model. The first three levels in his model correspond closely to Kirkpatrick's model. However, the final level is split into two: organization and ultimate value. The five level model is therefore:

- *Level 1: Reactions*
- *Level 2: Learning*
- *Level 3: Job behaviour*
- *Level 4: Organization* – the effects on the organization, from participant's job to performance changes
- *Level 5: Ultimate value* – the financial effects, both on the organization and the economy.

d) Guskey's critical levels

Thomas Guskey (2002) has also elaborated Kirkpatrick's 4 levels into 5: his levels may be of relevance as he had "students" and educational environments in mind (see the University of Minnesota website "Education Minnesota"<sup>27</sup>):

- *Level 1: Participant reaction*
  - Purpose: to gauge the participants' reactions about information and basic human needs
  - Technique: usually a questionnaire
  - Key questions: was your time well spent? Was the presenter knowledgeable?
- *Level 2: Participant learning*
  - Purpose: examine participants' level of attained learning
  - Technique: test, simulation, personal reflection, full-scale demonstration. Key question: did participants learn what was intended?
- *Level 3: Organizational support and learning*
  - Purpose: analyze organizational support for skills gained in staff development
  - Technique: minutes of district meetings, questionnaires, structured interviews or unobtrusive observations
  - Key questions: were problems addressed quickly and efficiently? were sufficient resources made available, including time for reflection?
- *Level 4: Participant use of new knowledge and skills*
  - Purpose: determine whether participants are using what they learned and using it well

---

<sup>25</sup> See listings published and analyzed by the UK Institute for Employment Studies.

<sup>26</sup> See A.C. "Hamblin's Evaluation and Control of Training", 1974.

<sup>27</sup> [www.educationminnesota.org/profdev/tall/Pages/5levels.aspx](http://www.educationminnesota.org/profdev/tall/Pages/5levels.aspx).

- Technique: questionnaires, structured interviews, oral or written personal reflections, examination of journals or portfolio, or direct observation
- Key question: are participants implementing their skills and to what degree?
- *Level 5: Student learning outcomes*
  - Purpose: analyze the correlating student learning objectives
  - Technique: classroom grades, tests, direct observation
  - Key question: did student show improvement in academic, behaviour or other areas?

e) Indiana University taxonomy

Indiana University developed an evaluation taxonomy-based on six strata, which were not intended to be a hierarchy of importance. The first and last strata provide additions to Kirkpatrick's framework:<sup>28</sup>

- *Stratum 1*: activity accounting – which examines training volume and level per participant
- *Stratum 2*: participant reactions
- *Stratum 3*: participant learning
- *Stratum 4*: transfer of training
- *Stratum 5*: business impact
- *Stratum 6*: social impact

The sixth stratum examines the impact of changed performance on society, and as such is similar to Hamblin's ultimate value.

f) Industrial Society stages

The Industrial Society (now the Work Foundation) developed a six stage circular model which starts with a planning phase. The stages are:

- *Stage 1*: identify the business need
- *Stage 2*: define the development objectives
- *Stage 3*: design the learning process.
- *Stage 4*: experience the learning process
- *Stage 5*: use and reinforce the learning
- *Stage 6*: judge the benefits to the organization (quality measures, customer satisfaction and financial benefits provide the main measures at this level).

The Industrial Society differentiated between stages 3 and 4 which aim to validate the training, and stages 5 and 6 which aim to evaluate it. True evaluation needs to take place long before and after training has taken place and the process of identifying the business need is an essential component of the evaluation model.<sup>29</sup>

---

<sup>28</sup> See "Designing Instructional Systems", in the ASTD Handbook, 1996.

<sup>29</sup> See Michael Armstrong "Industrial Society", 1998.

g) Kearns and Miller KPMT model<sup>30</sup>

Kearns and Miller's KPMT model has many similarities to Phillips work. They argue that clear objectives are an essential component of a training evaluation model. Where they differ is in their aim to provide a sort of toolkit to help evaluators work through the process of identifying bottom-line objectives by means of questioning techniques, evaluating existing training, and using process mapping to identify the added value to organizations.

They argue that training can only bring added value to organizations if the business is not performing effectively or there is a market opportunity which can be exploited. To identify bottom line benefits, pre-training measurements need to be in place. Only where the training is to bring someone up to the standards of the job is this not necessary.

The four-stage KPMT model starts at the beginning of the training cycle by identifying the business need rather than the training need. The emphasis is on clarifying objectives from a business perspective rather than that of the trainees. Despite this, the evaluation levels look very similar to Kirkpatrick's:

- Reaction to training and development
- Learning
- Transfer to the workplace/behaviour
- Bottom line added value, measured in relation to the base level measures taken

Where Kearns and Miller differ from some of the other models is in their belief that return on investment can only be looked at in hard terms. They state that if a business objective cannot be cited as a basis for designing training and development, then no training and development should be offered.

h) Nine outcomes model<sup>31</sup>

Also worth mentioning is the "Nine Outcomes" model which aims to measure whether training has been successful. In identifying the 9 outcomes, Donovan and Townsend pose 9 questions with the training participants in mind:

- *Reaction* to training – did they like it?
- *Satisfaction* with the organization of a training event (facilities, logistics, meals, etc.)
- *Knowledge* acquisition – did they learn anything?
- *Skills* improvement – can they do something new or better?
- *Attitude* shift – have they changed their opinions about something?
- *Behaviour* change – have they changed their way of doing things following the training?
- *Results* – how did the training impact on the organization's key success factors?

---

<sup>30</sup> See P. Kearns and T. Miller "Measuring the impact of training and development of the bottom line", 1997.

<sup>31</sup> See Donovan and Townsend "Training Evaluation Pocketbook", 2004.

- *Return on investment* – to what extent did the training give back more than it cost?
- *Psychological capital* – how did the training affect corporate image?

All 4 of Kirkpatrick's levels, incidentally, are included among the 9 outcomes.

i) Organizational elements model<sup>32</sup>

Kaufman and Keller (1994) argue that Kirkpatrick's model was intended for evaluating training, and that as organizations now seek to evaluate other types of development events, the framework needs to be modified. They expanded Kirkpatrick's model to include societal contribution as an evaluation criteria. They argue that manufacturing organizations in particular are increasingly being called to account for societal consequences such as pollution and safety.

The model also included some additions at the other levels, such as the inclusion of needs assessment and planning in the evaluation, an examination of the desired or expected results, and a review of the availability and quality of resources. They contend that evaluation at all levels should be planned and designed prior to the implementation of any intervention.

With the additional help of Watkins in 1995, the team reclassified the criterion in their model into the following six levels:

- *Level 1: Input* – similar to Kirkpatrick's reaction level, but has been expanded to include the role, usefulness, appropriateness and contributions of the methods and resources used;
- *Level 2: Process* – this level also has similarities to the reaction level, but is expanded to include an analysis of whether the intervention was implemented properly in terms of achieving its objectives;
- *Level 3: Micro (acquisition)* – this is similar to the learning level and examines individual as well as small-group mastery and competence;
- *Level 4: Micro (performance)* – links closely to the behaviour level and examines the utilisation of skills and knowledge. The focus is on application rather than transfer of skills and knowledge;
- *Level 5: Macro* – relates to the results level and examines organizational contributions and payoffs;
- *Level 6: Mega* – an additional level which looks at societal outcomes.

They argue that costs can be examined at each stage, from efficiency measures at the input level to utility costs at the highest level.

j) Contemporary ROI models<sup>33</sup>

A range of "contemporary" models/methods for assessing outcomes have been developed, elaborating on Phillips ROI in an effort to somehow better evaluate the ROI. On the face of it these methods have little relevance to public sector

---

<sup>32</sup> See Kaufman, Keller, and Watkins "What Works and What Doesn't, Evaluation Beyond Kirkpatrick", 1995.

<sup>33</sup> See <http://scholar.lib.vt.edu/ejournals/JITE/v41n3/brauchle.html>.

international organizations but are identified below if only to give an idea of the myriad of assessment methods available:

Methodology
Benefit/cost ratio
Payback period
Return on true value of dollars
Present value of dollars and future value of dollars
Utility analysis
360-degree feedback
Performance teams satisfaction
Balanced scorecard
HRD benefit forecasting

ii. Systems-based

Among these the CIPP, IPO and TVS models are perhaps the best known, though the distinction between “goal-based” and “systems-based” is sometimes ambiguous.

a) CIPP (Context, Input, Process, Product)

The CIPP model was developed by Daniel Stufflebeam from 1971 onwards (his latest work is “Evaluation Theory, Models, and Applications”, 2007). It distinguishes four types of evaluation (which we have encountered tangentially in earlier pages):

- *Context evaluation* – which helps in planning and developing objectives
- *Input evaluation* – which helps to determine the design by examining capability, resources and different strategies
- *Process evaluation* – which helps to control the operations by providing on-going feedback
- *Product evaluation* – which helps to judge and react to the programme attainments in terms of outputs and outcomes.

Corresponding to the letters in the acronym CIPP, this model’s core parts are context, input, process, and product evaluation. In general, these four parts of an evaluation respectively ask. What needs to be done? How should it be done? Is it being done? Did it succeed?

In this checklist, the “Did it succeed?” or product evaluation part is divided into impact, effectiveness, sustainability, and transportability evaluations. Respectively, these four product evaluation subparts ask. Were the right beneficiaries reached? Were their needs met? Were the gains for the

beneficiaries sustained? Did the processes that produced the gains prove transportable and adaptable for effective use in other settings?

As we will see later the “subparts” of the “product evaluation”, i.e. impact, effectiveness and sustainability, are also included in the World Bank’s Independent Evaluation Group evaluation framework and may be of particular relevance to public sector international organizations.

b) IPO (Input, Process, Output)<sup>34</sup>

Bushnell developed the IPO model (input, process, output) which focuses more on the inputs to training. The IPO model is used by IBM and helps to monitor employee progress by setting performance indicators at each stage. The stages are:

- *Input* – such as the instructor experience, trainee qualifications, resources
- *Process* – the plan, design, development and delivery of the training
- *Outputs* – the trainees reactions, knowledge and skills gained and improved job performance
- *Outcomes* – profits, customer satisfaction and productivity.

c) TVS (Training Valuation System)<sup>35</sup>

Fitz-enz (1994) developed a Training Valuation System (TVS) which is a four-step process similar to Kirkpatrick’s framework at steps 3 and 4 but has been categorized as “system-based”:

- *Step 1: Situation analysis* – this is similar to an in-depth training analysis. Like Kearns and Miller, he suggests that the manager’s answers are continuously probed until some visible, tangible outcome is revealed and that the questions initially focus on the work process rather than the training;
- *Step 2: Intervention* – this involves diagnosing the problem and designing the training;
- *Step 3: Impact* – this examines the variables that impact on performance
- *Step 4: Value* – this step places a monetary worth on the changed performance.

d) Pulley’s responsive evaluation model<sup>36</sup>

Another system-based evaluation model focuses on the purpose of evaluation, the “responsive evaluation” model developed by Pulley (1994).

Responsive evaluation is a tool for communicating evaluation results more effectively by tailoring it to the needs of the decision-makers. Pulley argues

---

<sup>34</sup> See D.S. Bushnell, “Input, Process, Output: a Model for Evaluating Training” 1990.

<sup>35</sup> See J. Fitz-enz “Yes, you can weigh training value”, 1994.

<sup>36</sup> See M.L. Pulley “Navigating the Evaluation Rapids”, 1994.

that the objective of the evaluation should be to provide evidence so that key decision-makers can determine what they want to know about the programme. The stages involved are:

- Identify the decision-makers so as to ascertain who will be using the information and what their stake in it is;
- Identify the information needs of the decision-makers – what do they need to know and how will it influence their decisions?
- Systematically collect both quantitative and qualitative data. Pulley argues that the qualitative data is normally relayed in the form of stories or anecdotes and gives life to the numbers;
- Translate the data into meaningful information
- Involve and inform decision-makers on an on-going basis.

e) E-Learning models

More recently, a range of system-based models have been elaborated for evaluating “new technology delivery” – such as online learning and e-Learning, as part of a portfolio of training options available to human resources managers less interested in instruction-led training (see for instance Pollard Hillage “Exploring e-Learning”, 2001).

One of many examples is the “Continuous Evaluation of Training Systems Based on Virtual Reality”. These models often seem to have been specifically adapted to evaluate technical scientific achievements.

f) Duignan’s framework for outcomes systems

Paul Duignan, to whom reference has been made in earlier pages, is also the main designer of a very complex system which develops intervention or programme logics using the Outcome Hierarchies diagramming approach. Intervention logics set out the connections between the outcomes an individual, organization, or group of organizations are trying to achieve and the steps, stages or intermediate outcomes which are needed to achieve this. It is a complex and highly technical approach which culminates in the identification of 7 high-level outcome attribution evaluation designs<sup>37</sup>. Duignan’s approach is also referred to as the OIIWA Systematic Outcomes Analysis, OIIWA standing for Outcomes Is It Working Analysis<sup>38</sup>. On the face of it the approach, which makes for tough reading, may seem farfetched for the UN agencies; of interest is, however, that the Duignan’s methodology (REMLogic) has been applied in 2005 in evaluating IMF surveillance activity. More on this later.

---

<sup>37</sup> See his “Principles of Outcome Hierarchies: Contribution Towards a General Conceptual Framework for Outcomes Systems (Outcomes Theory)”, 2004.

<sup>38</sup> See [www.oiiwa.org](http://www.oiiwa.org).

g) United Way of America outcomes evaluation

Outcomes evaluation (which is also an evaluation type) can be less complex than above. The UWA describes a step by step plan for this kind of evaluation, in its 1966 “Measuring Program Outcomes: A Practical Approach”. Of notable interest is the identification of differences between:

- *Outputs* – which indicate little about changes in trainees and are usually just numbers;
- *Outcomes* – which indicate real changes in trainees;
- *Outcome targets* – which specify how much of one’s outcome one hopes to achieve;
- *Outcome indicators* – which suggest progress towards the outcome targets.

h) Charities Evaluation Service outcomes triangle<sup>39</sup>

The CES has applied an outcomes approach for a range of UK governmental entities, notably the London Housing Authority. The CES defines outcomes as the effects of activities, the changes, benefits or learning that occur as a result of work carried out. Outcomes are neither outputs nor user satisfaction. The former are the detailed activities, services and products of organizations; the latter usually involves asking clients/trainees what they think about different aspects of the services etc. provided. While both are important they are not outcomes. The CES applies a triangular "Outcomes Learning Cycle" to help in the process of clarifying and measuring outputs, in analyzing what can be learned from the outcomes achieved, and planning/implementing changes as a result of such learning. Additionally, it links outcomes to aims which need to be clearly defined. The CES differentiates between aims and objectives, aims being the changes one hopes to achieve as a result of one's work, while objectives are the activities undertaken and the services offered to bring about the changes (this differentiation is clearly not accepted by all evaluation specialists). Details of how CES has applied its method can notably be found in its "Managing Outcome", 2003.

iii. Additional evaluation methods, including Contribution Analysis

The models/methods noted above do by no means exhaust the field. Several other methods or approaches exist, some of which may have less relevance for public sector international organizations. A few of these are still identified below:

Approaches
Dixon's six steps
Brinkerhoff's six stages
Bramley's goal-based
Wade's high impact
Shapiro's matrix

<sup>39</sup> See inter alia [www.homelessoutcomes.org.uk/Outcomes\\_approach.aspx?print=Y](http://www.homelessoutcomes.org.uk/Outcomes_approach.aspx?print=Y).

Pershing's perspectives
Warr, Bird and Rackham's CIRO (Context, Input, Reaction, Outcome)
Preskill and Torres' Evaluative Inquiry
Kraiger's Learning Outcomes
Kaplan and Norton's Balanced Scorecard

The International Board for Standards, Training , Performance and Instruction (the IBSTPI) has published a 12 page succinct and useful summary of approaches not mentioned in the text above, which may be of interest  
[http://media.wiley.com/product\\_data/excerpt/91/07879959/0787995991.pdf](http://media.wiley.com/product_data/excerpt/91/07879959/0787995991.pdf).

Of particular relevance is no doubt Contribution Analysis. A relatively recent approach to evaluation in international development it has inter alia successfully been applied by Australia's AID. It centers on a system of outcome based monitoring and evaluation and could equally well be included under item 8.ii above. For general references on this approach see inter alia [www.developmentgateway.com.au/jahia/Jahia/pid/4125](http://www.developmentgateway.com.au/jahia/Jahia/pid/4125) as well as documents at [www.aes.asn.au/conferences/2006/papers/022%20Fiona%20Kotvojs.pdf](http://www.aes.asn.au/conferences/2006/papers/022%20Fiona%20Kotvojs.pdf).

Contribution analysis can be linked to “theory-based evaluation” (see in this context the American Development Bank efforts at the NONIE website).

Contribution Analysis is an innovative approach which may consider dimensions sometimes overlooked by more traditional methods; it can however be used in conjunction with other models and requires considerable intellectual precision. It is our good fortune that Professor Rotem, highly knowledgeable in the field, analyses the approach in depth as a potential evaluation framework for UN system organizations in his report.

## 9. Critique of the Kirkpatrick model

In particular there have been questions with respect to the levels of reaction and learning. Before assessing reactions some critics call for a broad analysis of the organizational context, its values, practises and current situation. Also, they call for a more explicit focus on the needs of the organization and how these tie to the development of objectives and the design of the most appropriate solution. Research undertaken has shown that, just because people liked a course, it does not necessarily mean they learned anything. In some cases it appears that the more the trainees liked a course, the less they learned.<sup>40</sup>

Within the results level there are suggestions the benefits to the organization should be made more explicit and focus on monetary values such as ROI. There may also be a need for evaluating beyond the organization by examining the effects on the economy and the societal consequences.

<sup>40</sup> See Alliger and Janak, “Kirkpatrick’s levels of training criteria: thirty years later”.

Additionally, there is a need to identify the reasons for the evaluation; the tools and techniques employed will alter depending on why the evaluation is taking place and who is for it. The evaluation strategy must be tailored to the audience within the organization rather than putting measures in place just for the sake of it.

For instance, the Indiana University website cited below identifies 7 specific limitations of the Kirkpatrick model:<sup>41</sup>

- Not situation driven
- Not programme specific
- It depends on contextual needs
- You need to specify unit of analysis (groups or teams, individual, organization)
- Misleading – Levels are different perspectives not a hierarchy
- Level 1 can only provide participant reaction and is subjective
- Most companies do level 1 or level 2 evaluations which only provide limited information.

E.F. Holton (“The flawed four-level evaluation model”, 1996) is one of the main critics. To him the levels form a “taxonomy” of outcomes rather than a model, due mostly to the assumption of causal relationships between the levels that are not empirically tested. Holton also argues that no evaluation can be validated without measuring and accounting for the intervening variables that effect learning and transfer processes. Kirkpatrick provides a model for thinking about how to evaluate but the model does little to inform what to evaluate and how to link the results to strategy.

Holton and other critics specifically note that Kirkpatrick’s model:<sup>42</sup>

- Implies a hierarchy of values related to the different values, with organizational performance (result) being seen as more important than reaction, etc.;
- Assumes that the levels are each associated with the previous and next levels; this causal relationship, it is argued, has not always been established by research;
- Is too “simple” and fails to take account of the various intervening variables affecting learning and transfer;
- Implies correlation between learner reactions and measures of learning and subsequent measures of changed behaviour – but we know now that “satisfaction” is not necessarily related to good learning and changed behaviour (see also above);
- Implies that performance during training is a prediction of post-training performance;
- Ignores the frequent failure of training to transfer into the workplace (due to the range of organizational factors which may inhibit success);
- Ignores – when it comes to the level of “result” – the inherent difficulty on linking soft skills training to hard results.

---

<sup>41</sup> See [www.Indiana.edu/~jstr561/knuth06sum/unit1g.shtml](http://www.Indiana.edu/~jstr561/knuth06sum/unit1g.shtml).

<sup>42</sup> See also G. Alliger and E. Janak “Kirkpatrick’s levels of training: 30 years later”, 1989.

## 10. Feasibility of training impact assessment/some success stories

An excellent paper on the whole subject has recently been compiled by Deloitte Insight Economics, i.e. “Impact Monitoring and Evaluation Framework”, June 2007, referred to on p.4. In addition to proposing an impact evaluation model and a framework for impact assessment, the report adapts the proposed framework to agriculture and rural based technology, environment, manufacturing technology, information and communication technology, mining and energy, and medical science and technology. Moreover, the report identifies the “types of benefits” generated in these sectors and provides an extensive “bibliography” of evaluation surveys carried out in these same sectors in Australia; the report is methodologically quite complex and does not per se identify “success stories”.

Though there is much evidence to suggest the contrary some organizations have reported a measure of “success” in determining the results of training, at least in the sense that such results did or did not prompt change in the training programmes based on assessment criteria. For instance, there have in recent years been several evaluations of executive development programmes conducted in and by the Canadian federal government. These evaluations broadly applied only level 1 and 2 of the Kirkpatrick’s model (yes, the Canadian authorities are apparently adherents to it), but confirmed the effectiveness of the programmes, within these levels, programmes which were subsequently “fine-tuned” and modified as a consequence of the evaluations. It was recognized that evaluations at higher levels (3 and 4) are difficult in the public sector; even so, attempts at higher-level evaluation have been made with respect to the Accelerated Executive Development Program.<sup>43</sup>

The website of the US Department of Energy’s Energy Efficiency and Renewable Energy unit provides concrete and detailed examples of “evaluation successes”; the EERE seems to use a mixture of evaluation methods and gives us no specific guidance as to what evaluation framework is the “best”.<sup>44</sup>

The external evaluation reports (2000-2004) by the US Center for Disease Control also noted “success” in the evaluation of Vietnam’s National Tuberculosis Program (NTP) with regard to the annual Management for International Public Health (MIPH) course, as concerned.<sup>45</sup>

- Determining the effects of the MIPH course on graduates from Vietnam;
- Assessing the quality and sustainability of the management course in Vietnam developed by MIPH graduates from NTP, in collaboration with MIPH graduates from the Hanoi School of Public Health (HSPH);
- Analyzing the impact of the NTP management course on learners and their organizations in Vietnam.

WMO, through its National Meteorological and Hydrological Services (NMHSs) reported “positive” evaluation results and the achievement of a Media Training Course objectives in an impact evaluation, dated February 2006. A note of caution was included, however, in the third conclusion of the brief report, i.e.:<sup>46</sup>

---

<sup>43</sup> See reports of the Treasury Board of Canada at [www.tbs-sct.gc.ca/eval/pubs/eet-efcs/eet-efcs\\_e.asp](http://www.tbs-sct.gc.ca/eval/pubs/eet-efcs/eet-efcs_e.asp).

<sup>44</sup> See [www.eere.energy.gov/ba/pba/program\\_evaluation/success\\_stories.html](http://www.eere.energy.gov/ba/pba/program_evaluation/success_stories.html).

<sup>45</sup> See [www.cdc.gov/smdp/docs/VNEval.pdf](http://www.cdc.gov/smdp/docs/VNEval.pdf)

<sup>46</sup> See [www.wmo.ch/pages/prog/amp/pwsp/documents/MediaCourseEvaluation.pdf](http://www.wmo.ch/pages/prog/amp/pwsp/documents/MediaCourseEvaluation.pdf).

- The objective of the course, which was to equip participants with improved communication and presentation skills was achieved;
- Passing of knowledge to colleagues in NMHSs by the participants vindicates the high value the participants attached to the training event, and is providing a multiplier effect in spreading acquisition of skills in NMHSs;
- While improvement to radio and television presentations will definitely benefit from this training event, no participant indicated how they will apply their skills to presentations on the Internet or on the print media. These could be areas that need to be considered for future events.

The International Council of Nurses conducted an evaluation study of “Leadership for Change” in 2001-02 focusing on impact and sustainability. The results were deemed “excellent” as the “findings indicate significant positive results, especially at the level of individual development”.<sup>47</sup> The ICN report as the WMO report cited above do not, unlike the Canadian government, specify any particular evaluation model applied but the ICN indicated that the evaluation report looks into issues pertaining to relevance and sustainability which of course form part of the World Bank/IEG approach.

Concrete examples of “successful” impact assessment have been reported in a wide range of countries (Australia, Brazil, Israel, Japan, Venezuela, etc) but the scope of this paper does not permit an exhaustive review. However, it does appear that these evaluations have rarely gone beyond the parameters of levels 1 and 2 of Kirkpatrick’s model. They nevertheless give support to the assumption that impact assessment of some kind is feasible in the presence of political will and a genuine evaluation framework.

The World Bank / IEG website ([www.worldbank.org/ieg/intro](http://www.worldbank.org/ieg/intro)) identifies the considerable number of evaluation studies undertaken by the Bank throughout the globe and is of great interest though it does not per se provide a “scoreboard” in terms of failures and successes. Of relevance, though not specifically a case study of training evaluation, is also Judith Tandler’s “Good Governance in the Tropics” (1997) which examines the factors which permitted a state in northeast Brazil to move from poor government performance to internationally acclaimed rural development service at local level.

In an earlier section reference was made to the complex “Outcomes Hierarchies” model developed by Paul Duignan. In fact, the IMF’s Independent Evaluation Office (without endorsing the findings) in 2005 published Mr Duignan’s lengthy paper on “Strategy Design in Evaluating IMF Surveillance Activity” in which he (and the co-author Nils Bjorksten) apply the REMLogic to “help improve the overall effectiveness of surveillance”. REMLogic stands for Research Monitoring Intervention Logic (Outcomes Theory). In collaboration with PricewaterhouseCoopers Mr Guignan has also applied the REMLogic in an evaluation exercise in 2005 for New Zealand’s Department of Building and Housing; he has been active in a range of such analyses<sup>48</sup>, including an evaluation exercise (with Carolyn Lane) for New Zealand’s publicly financed social programme sector, noting that “successful” evaluation required an “Evaluation Culture” which measures outcomes rather than, as in the past, outputs.<sup>49</sup>

The success or otherwise of these initiatives is not known.

<sup>47</sup> See [www.icn.ch/LFCoutcomes.htm](http://www.icn.ch/LFCoutcomes.htm).

<sup>48</sup> See [www.strategievaluation.info](http://www.strategievaluation.info).

<sup>49</sup> See [www.parkerduignan.com/se/documents/103f.html](http://www.parkerduignan.com/se/documents/103f.html).

A significant UN entity contribution to the outcomes method, which provides concrete examples of a) outcomes, b) outputs, and c) "indicators" (needed to help describe how the intended outputs will be measured) to several sectors has been made by the United Nations Development Group. The sectors include poverty reduction, good governance, basic services, gender equality and women's rights, and refer to Moldova, Kyrgyzstan, Afghanistan, and Timor Leste. The UNDG report is not a listing of "success stories" per se but its highly interesting methodology makes it recommended reading.<sup>50</sup> Of interest is also the UNDG "checklist for validating outputs/guidance note on indicators". The indicators used are very specific and detailed, since they are intended to force clarification of what is meant by output and are used for monitoring, evaluating, and reporting on achievement.

Not much attention has been devoted in these pages to the subject of "indicators", if only to respect a measure of conciseness. Indicators would, however, be of priority concern to the Task Force. In this context reference is also made to a WHO Regional Office document entitled "Putting the evaluation indicators and evaluation system into practical use".<sup>51</sup> General sources indicators can be found inter alia at:

[http://ec.europa.eu/europeaid/evaluation/methodology/methods/mth\\_ind\\_en.htm](http://ec.europa.eu/europeaid/evaluation/methodology/methods/mth_ind_en.htm) and at [www.uap.vt.edu/checkyoursuccess/workshop.html](http://www.uap.vt.edu/checkyoursuccess/workshop.html); the latter is a "guide for creating your own indicators" and is useful for a step by step approach on developing good indicators.

### III. Measures to support an evaluation framework

#### 1. Current state of fellowships and changing expectations

The generally pessimistic view of many agencies as to the volume of fellowship activities is somewhat contradicted by information received from the Inter-Agency Procurement Services Office. JIU reports indicate that "while the share of fellowships awarded by agencies remained stable, fellowships awarded under NEX (national execution) have almost doubled". There seems nevertheless to have occurred a further downward trend in agency awards, certainly as concerns traditional, individual, project oriented fellowships of fairly long duration (six months or more) and, with some exceptions, also as concerns study tours. In a number of cases the fellowships services have been significantly affected in terms of staff and other resources, sometimes resulting in serious morale problems. The situation is rendered complex by the apparent survivability of fellowships in several agencies, especially outside the UN system. Within the system, new formulas have been found to replace dwindling UNDP funds, whether in terms of unilateral trust funds, co-sponsored schemes, or multilateral/bilateral course sponsors.

Another question relates to the trends which have affected the fellowship programmes. At least ten can be noted (others could surely be added):

- Decline in long term fellowship awards and increased emphasis on short and medium term training: the latter are often outside the realm of "fellowships";

---

<sup>50</sup> The UNDG draft document, as is true for other UNDG reports on outputs, outcomes, and performance indicators, can be researched from the main website <http://www.undg.org/>. See also [www.undg.org/archive\\_docs/9284-Technical\\_brief\\_-\\_indicators.doc](http://www.undg.org/archive_docs/9284-Technical_brief_-_indicators.doc)

<sup>51</sup> See <http://www.wpro.who.int/> - look for file CHN 2007 Dec Evaluation Indicators System.doc

- Decline in UNDP resources available for agency implementation, and an increase in nationally executed programmes which have been further encouraged by modern technology (Internet, etc.);
- Increase in placements in developing countries and, as a corollary, the increased regionalization of placements;
- Increase in the percentage of awards granted to women;
- Increase for some agencies in funding from regular budgets and/or alternative sources;
- Increased use of distance learning methods, as a consequence of modern information technology, sometimes to the detriment of traditional training methods;
- Increased resort to in-country and on-the-job training in the home country (considered more effective and less expensive than studies abroad);
- Aggressive competition from UNDP which has sometimes monopolized certain segments of the fellowship market with its advantage of the UNDP local and regional network;
- The development of training expertise by the World Bank and a range of bilateral institutions which have overlapped UN agency mandates and can offer similar training with greater resources;
- An apparent decline in political interest in fellowships.

Concerns over the state of fellowships and the sometimes negative trends which seem to have affected the programmes have dramatically increased demand for accountability, notably through evaluation and impact assessment. Regardless of approach or model chosen it is also evident from literature on the subject that organizational expectations for training have shifted dramatically. The fact is that the nature of training itself is undergoing a transformation. Trainers no longer hold the privileged position of “all knowing” content expert. Groups being trained often contain individuals with more depth of knowledge about, more experience applying, or more time to access current knowledge on the subject of the training. The training professionals thus become facilitators of learning and guides to available knowledge instead of content experts who bring “the info” into the training room with them. Trainers no longer “own” the knowledge. Instead, they synthesize and provide resources to clients who also have access to the knowledge.

As training has moved from satisfying trainers to improving organizational performance, the definition of customer has broadened. Trainees themselves are still among the “customers” of training – and the trainee’s evaluations are important sources of feedback for continuous improvement and quality – but the trainee’s organizational unit and the organization as a whole are now part of the client system. Training is performed to solve the problems of the unit and have a positive impact on the organization.

## 2. Fellowships evaluation in selected bilateral and multilateral organizations

The question is whether, and how, fellowships are being evaluated by UN system and other agencies. Again, there is no uniform answer for the situation varies considerably. However, while evaluation exercises have been carried out on a non-systematic basis and apparently not within an evaluation “framework”, these, as far as has been determined in

the course of the SFO meetings, do not truly measure up to the standards of “impact assessment”. The agencies and entities which are training oriented fare best and, as a consequence of the 1998 JIU report, several agencies have undertaken relatively limited evaluations of their fellowships programmes. Internal evaluations of the fellowships component, however, are normally undertaken only in the global context of the evaluation of projects and programmes, and it is noted that fellowships remains the smallest component as compared to experts/consultants and equipment. Many agencies have a central evaluation unit responsible for such project/programme evaluation but fellowships have not been a main focus for assessment.

Joint evaluation exercises with a donor/sponsor can be useful. External evaluations can yield interesting results, leading to remedial action by the agency concerned. The question has also been raised whether it makes sense to evaluate fellowships in the narrow sense as opposed to wider training activities. As concerns resources these are generally pitiful and do not exceed 1 per cent of adjusted programmes. By and large, fellowships seem inadequately evaluated.

In this context some concepts<sup>52</sup> applied by various agencies and entities may be retained as particularly relevant for fellowships evaluations, i.e.:

*i. Evaluation “yardsticks” (United Nations Institute for Training and Research)*

1. Framework of the training programme (mandate, rationale, objectives, etc.) – if the fellowships component is not adequately defined from the outset it will be proportionately more difficult to attempt an assessment of its impact;
2. Effectiveness of programme – the extent to which the programme achieved its objectives and reached its target group;
3. Overall efficiency of programme – the extent to which the results achieved by the programme still justify the costs incurred;
4. Impact of training activities in target countries – the extent to which the general objectives set for the programme have been translated into general development in the countries;
5. Proposals for future training activities – based on the assessment, evaluators should develop recommendations for the future focus of the programme;
6. Proposals for sustainability – the likelihood that the programme benefits will be maintained locally after withdrawal of external support and funding.

*ii. The four “levels” of evaluation (ILO International Training Centre, Turin)*

- 1) Evaluation of participants’ satisfaction;
- 2) Evaluation of individual progress (knowledge and skills required);
- 3) Evaluation of impact on the working behaviour of participants (changes as a result of learning/training);
- 4) Evaluation of impact at the institutional and organizational level.

*iii. Training “benefits” (British Council)*

---

<sup>52</sup> These “concepts” were identified some years back and may well have been changed/modified/expanded or even eliminated since. In any event they do not constitute models per se (except perhaps for the World Bank (IEP)).

- 1) Benefits to the individual – the impact of the training in terms of performance and qualifications required;
- 2) Benefits to the work group – benefits which trainees are able to pass on to colleagues in their work group;
- 3) Benefits to the organization – the wider benefits to the organization with examples of impact at this level.

iv. *The three “D”s and “E”s of evaluation (UNESCO)*

- 1) The capacity to detect emerging problems;
- 2) The capacity to design an international action;
- 3) The capacity to develop the process;
- 4) The need for economy;
- 5) The need for efficiency;
- 6) The need for effectiveness.

v. *The three “categories” of evaluation (FAO)*

- 1) The modality, in terms of the evaluator’s institutional relationship to the activities under evaluation:
  - a) auto-evaluation
  - b) internal
  - c) external
- 2) The timing of the evaluation:
  - a) ex-ante
  - b) on-going/mid-term
  - c) terminal or final
  - d) ex-post
- 3) The purpose of evaluation:
  - a) formation
  - b) summative
  - c) impact and sustainability

### 3. The objectives-based World Bank (Independent Evaluation Group) approach

First, it must be admitted that specialist literature generally sees the Kirkpatrick’s model and/or of its variants of the goal-based variety as the most widely used in training evaluation, if only for its “elegant simplicity”.<sup>53</sup> System-based models such as CIPP and IPO have certainly also their advocates. However, the ASTD reveals that 67% of organizations conducting evaluation use the Kirkpatrick’s model.

For public sector international organizations the World Bank / IEG approach may have advantages over the Kirkpatrick and/or similar models if it is deemed that it enhances accountability by better focusing attention on (1) the extent to which objectives agreed with the governing bodies of the agencies have in fact been achieved, and (2) the promotion of efficiency by better relating the use of scarce resources to the

<sup>53</sup> See “Evaluation of Training” by Stone and Watson.

accomplishment of specific objectives. Most of all perhaps, the World Bank “environment”, while surely unique, may not be all dissimilar to that faced by the UN and its agencies. A number of bilateral development agencies, including that of Japan, have adopted the IEG approach.

The World Bank (IEG) approach is identified below:<sup>54</sup>

i. Outcome

The IEG evaluates outcome by considering three factors:

- *Relevance* of the intervention’s objectives in relation to country needs and institutional priorities;
- *Efficacy*, i.e. the extent to which the developmental objectives have been (or are expected to be) achieved; and
- *Efficiency*, i.e. the extent to which the objectives have been (or are expected to be achieved) without using more resources than necessary.

ii. Sustainability

The sustainability of the project measures the likelihood that its estimated net benefits will be maintained or exceeded over the life of the project. Sustainability reflects the resilient risks of a project as measured by the likelihood that its estimated net benefits will be retained or exceeded over the project’s intended useful life.

iii. Institutional development impact

The institutional development impact measure evaluates the extent to which a project improves the ability of a country/region to make more efficient, equitable and sustainable use of human, financial and natural resources. The IEG evaluates each project’s success in fostering changes.

#### 4. The five requirements for fellowships evaluation

Another question relates to the requirements for fellowships evaluation. These may be pretty well known but are rarely met other than in ideal circumstances. Common sense though they are, they bear repetition if only to demonstrate how far removed the fellowships component is from the “purist” view of evaluation. In a somewhat simplistic vein evaluation of fellowships requires us to **smart**, i.e. to select fellows, to monitor them, to have local and regional allies for their supervision, to adequately retrieve information on and about fellows, and to maintain ties with fellows after the training.

Does any UN system agency fully meet these requirements? Is any agency/entity truly **smart**? Probably not, but improvements should surely be sought when possible and, in those instances where a modicum of the requirements cannot be met, the cessation of fellowship awards should be envisaged to avoid a corruption of this training component.

Admittedly, being “**smart**” will not suffice unless projects have been adequately designed from the perspective of the fellowships component. Prior to the actual selection process we must know the purpose of the fellowship. Assuming we know what we are trying to accomplish, it is then up to us to be **smart in terms of the five requirements**, i.e.:

---

<sup>54</sup> See [www.worldbank.org/ieg/ieg\\_approach.html](http://www.worldbank.org/ieg/ieg_approach.html).

i. Selection

All agencies have selection criteria but these serve little purpose if they are not adhered to (which too often seems to be the case), and if the implementer is not, at least to a degree, the selector. Certainly the agencies should assert their priority role in the selection process when it concerns projects funded by the system itself.

In too many instances there is an inadequate screening by the agencies of the nomination forms. We should have the courage to say “no” if faced with too demonstrably flawed candidates. We should ask the questions of “why” we need this training and “who” deserves to be trained. New forms of sponsored and co-sponsored fellowships have often removed these from the project context and in these instances the funding source has a stronger claim to the selection decision. For sponsored programmes it may not be unreasonable to insist on the donor/selector to take responsibility for evaluation, perhaps on a joint basis.

ii. Monitoring

All agencies require fellows to submit reports to permit the monitoring of their progress. The frequently poor response rate holds mostly true for agencies which are deficient also in terms of the other evaluation requirements. Perhaps greater use should be made of various incentives to encourage fellows to submit meaningful reports (whether in the form of certificates or withholding of payments). Reports from fellows about their training do not suffice, however. There is a need for comprehensive reports about fellows from the training institutions, and it is surprising that these are not universally required and obtained. There is also a need for reports from fellows well after their training and from their end/beneficiary users, to assess the impact of the training in the home work environment (and not only the increased knowledge of the fellows). Reports from the end user are admittedly difficult to obtain, unless the collegial/professional links between the implementing agency and the end user are very strong.

iii. Allies

Agencies which have a local/regional network of allies and partners are inherently advantaged since they can in principle count on their allies to guide and supervise the fellows during, and after, the training. Agencies/entities which lack allies will naturally be handicapped in their evaluation efforts. Those agencies may run “message-less” programmes, are unlikely to identify allies, precisely because of their non-specificity, and should probably abandon any pretence to run a training programme.

iv. Retrieval

The revolution in information and communications technology has brought new opportunities in fellowships management but not everyone has fully seized these opportunities. Huge quantities of reports from and about fellows lie on shelves, unexploited and essentially wasted. There is a need to create data bases from the information available so it can be retrieved when required.

v. Ties

We need ties with the fellows’ home and work environment if we hope to be able somehow to measure the impact of the training. If there is no follow-up to the monitoring of fellows during the actual training, we will never be able to assess the eventual contribution of fellowships to national capacity and institution building. It is

important to obtain reports from fellows sometime after the training but it may be even more imperative to engage the end/beneficiary user in a dialogue. A whole range of alumni activities could be envisaged. There are reasons to believe that the maintenance of such ties should be possible for any agency/entity which has a real message to transmit through its fellowships programme.

## 5. Key findings

A few observations or “findings” can perhaps be ventured at this stage:

- A. Despite a significant decrease of traditional individual fellowships, a few UN agencies/entities maintain vigorous fellowship programmes through new formulas, having found alternative funding sources to the UNDP.
- B. These agencies have been relatively successful because their fellowships partake of a UN system “message”, with a set of values which render them a bit distinct and not directly in competition with other, often financially more powerful sponsors.
- C. Agencies which have been unable to find alternative funding sources have jeopardized their fellowship programmes; it may be surmised that their programmes did not carry a specific UN system “message” and lost out to competition with other sponsors (for which reasons their disappearance may not be a cause of regret).
- D. Very few agencies/entities in the UN system meet the requirements for adequate fellowships evaluation and measures should be taken to remedy this situation, especially in an era of results-based budgeting.
- E. By and large the fellowships services do not engage in evaluation per se, or at best only peripherally. In some agencies there exist distinct evaluation units, outside of the fellowships services, but fellowships as such are rarely if ever the precise target of evaluation. Where relevant evaluation does take place it is focused on a relatively small portion of the total technical cooperation projects/programmes, and fellowships are evaluated as one, and the smallest, component of technical cooperation.
- F. The basic recommendation, with a view to remedial action, is that the UN system agencies endeavour to improve fellowships evaluation by being **smart**, i.e. by paying more attention to the need for:
  - a) selection of fellows;
  - b) monitoring of fellows;
  - c) allies to guide and supervise fellows;
  - d) retrieval of data on fellows;
  - e) ties with fellows and the end users (in the home work environment after the training).

The situation with bilateral and other multilateral institutions is quite complex, but by and large their fellowship programmes have retained a greater vitality than in the UN system, on a variety of grounds, but probably mostly linked to their greater control of the nomination and selection process.

- G. In the evaluation of fellowships the question is inter alia, and rightly so, posed as to their impact on capacity building. UN and other system agencies, once and if they have assessed the quality of the fellowships programme as being good, today pose a broad range of questions: (a) do fellowships bring about primarily an increase in individual knowledge? (b) does the increase of individual knowledge have an impact

- on the capacity of beneficiary institutions to induce desired change and improvement?  
(c) does institutional change/improvement also lead to sectoral change/improvement?  
(d) as concerns the UN system, is there such a thing as a UN system fellowship and a UN system fellow, somehow unique, with a clear message and values rendering them distinct from bilateral and/or private sector programmes?
- H. To sum up the debate on fellowships, their impact, and their evaluation, one may be tempted to adapt the saying that “writing comes more easily if you have something to say” into the proposition that “fellowships will have more impact if they carry a message” (linked to organizational objectives). The decrease of some UN fellowships programmes has many reasons but the absence of uniqueness and a distinct message may well be the major one.
- I. In the end it is worthwhile retaining only those fellowships which are demonstrably compatible with well defined organizational and developmental objectives and in respect to which an evaluation along the lines suggested above is feasible. Whatever activity cannot be evaluated is logically not impact measurable and is likely to be ineffective, if not detrimental, in promoting organizational objectives. As for study tours, there may be good reasons to consider a significant reduction even before completion of any formal evaluation exercise and to be honest in labelling them as political networking exercises.
- J. It is also high time to cease the quantifying of fellowship statistics in terms of numbers and man-hours, etc. What counts is the quality of the training provided under fellowships, their compliance with strictly defined organizational objectives and their impact measurability in terms of evaluation requirements. In this context the UN system agencies could consider spending more time and resources in identifying the right persons for fellowships, i.e. the enablers and achievers who may eventually come to “own” the development process in their country. The current nomination process too often disengages the organizations from the selection responsibility.
- K. The agencies should also consider increasing resources for evaluation, as have done many institutions concerned with training, notably the World Bank and the British Council. Evaluation does cost but “spending whatever limited funds are made available for a fellowship programme without having any reasonable indication of impact is a waste of much needed resources” (1998 JIU report). One is tempted to paraphrase Thoreau: “If you think evaluation is expensive, think of the cost of ignorance” (the great philosopher in fact referred to “education” but the parallels may not be so far fetched).
- L. Both substantive and fellowship services should assume greater responsibilities in the above selection/identification process; the often cumbersome, labour intensive and time consuming administrative procedures associated with traditional, individual fellowships and study tours cannot be suppressed overnight but should be gradually de-emphasized in favour of services linked to the more substantive preparatory, supervisory and follow-up elements of contemporary training components which seem to better support organizational and developmental objectives in many bilateral and multilateral agencies outside the UN system.
- M. The key to the effectiveness of fellowships, whatever the modality, appears to be the degree of control exercised over the selection process by the sponsors/implementers. Other factors obviously play a role but the experience of many institutions, inside and outside the UN system, multilateral or bilateral, identifies no magical formula but only the need for vigilance in the selection of “enablers and achievers”. For this reason

alone the current nomination process is inherently suspect and is in need of a much stronger element of screening and selection. Some bilateral institutions, like the US NIH, have apparently improved the nomination and selection process through the substitution of institutional for individual awards, with the concomitant establishment of collaborating centres of excellence.

N. In conclusion, the agencies may inter alia wish to consider the following:

- a) ensuring adequate resources for evaluation and initiating evaluation, with an adequate monitoring system, prior to commencing any fellowships process;
- b) ensuring that fellowship programmes are compatible with organizational objectives, adequately defined within a mission statement, and that impact findings be submitted to the governing bodies more systematically and regularly;
- c) ensuring greater control of the nomination process and greater involvement in all the phases of evaluation, selection, monitoring, liaison with allies, data retrieval and maintenance of ties with fellows;
- d) fellowships, which do not permit such evaluation, may be impact deficient and even detrimental to organizational objectives, and should be deleted;
- e) whatever their value, study tours are considered by many a misnomer and should be replaced by concepts such as visitorships, observation visits, or simply short term training; study tours, whatever the semantics, are often unfocused and tainted by an uncontrollable nomination process which can bring disrepute to the Organization; if so, they should be deleted and only scientifically and politically legitimate observation visits retained;
- f) more resources should be spent on ways and means of selecting the right persons for fellowships and short term training, i.e. the enablers and achievers who are likely to come to exercise decision-making in their beneficiary countries;
- g) the concept of institutional rather than individual awards should be borne in mind, with the establishment of partner institutions and collaboration with national centres of excellence;
- h) the benefit of regional training centres should be borne in mind, with the increasing resort to a range of in-country training exercises, while maintaining the element of international exposure.

#### IV. Conclusions and recommendations

1. For the UN and its agencies fellowships evaluation is rendered particularly difficult in that “fellows” are not trainees cum employees in the institutional sense. Such evaluation as has taken place has not been “impact conscious”. In this context “impact” has been defined by NONIE ([www.oecd.org/dataoecd/19/29/40104352.pdf](http://www.oecd.org/dataoecd/19/29/40104352.pdf)), the Network of Networks on Impact Evaluation, which includes UNEG; DAC; ECG, etc., as “positive and negative, primary and secondary long-term effects produced by a (development) intervention, directly or indirectly, intended or unintended” (for other technical definitions see the 40 page OECD/DAC “Glossary of Key Terms in Evaluation and Results Based Management” at: [www.oecd.org/dataoecd/29/21/2754804.pdf](http://www.oecd.org/dataoecd/29/21/2754804.pdf))
2. Despite continued scepticism over the prospects of evaluating the impact of fellowships, especially on a system-wide basis, among agencies with diverse objectives, different stakeholders, and varying management styles, a consensus seems

to have been reached, as inter alia evidenced by the recent UNDP regional workshops, that additional efforts to find common ground are indispensable for purposes of accountability. The UNEG Norms and Standards for Evaluation, though not focused on fellowships (too often ignored!), are noteworthy in this context (see Annex).

3. To render such efforts meaningful in terms of evaluating fellowships from the perspective of impact – not, as heretofore, merely the satisfaction of fellows – a system-wide evaluation framework is required. An evaluation framework should, ideally, specify various details of the evaluation process identified in the preceding pages. It can be defined formally as “a systematic accounting of the criteria considered and methodologies applied in determining the impact of measures on stakeholders and/or stakeholder interests” (see [www.losl.org/boardroom/glossary\\_e.php](http://www.losl.org/boardroom/glossary_e.php)). This is obviously no small business. It required almost 200 pages for the UK government to develop a methodology and evaluation framework in 2006 for its study on “Evaluating the Impact of England’s Regional Development Agencies” ( see inter alia documents at [www.berr.gov.uk/files/file21900.pdf](http://www.berr.gov.uk/files/file21900.pdf)):
4. Such a framework notably needs to be supported through both formative and summative evaluation (types) with some attention also provided to confirmative and meta evaluation types. In reality, however, one may doubt whether UN system agencies have the resources to engage in the latter two types, which are often even beyond the reach of wealthy corporations. Even a small effort in this direction would be welcome.
5. A wide range of evaluation models or methods exist, essentially goal-based and/or system-based. The Task Force will need to decide what model or combination of models/methods could be appropriate to support an evaluation framework. Kirkpatrick’s Four Levels, or one of its variants, has made the biggest inroad in the private and even the public sector but other models could be considered, especially the World Bank (IEG) approach which has the advantage of apparently proven merit in a public international organizational environment. More recent approaches, such as Contribution Analysis, should also be considered thoroughly.
6. Whatever the model or method chosen the foundations for training need to be laid before the training programme takes place, and the training objectives must be made more explicit to permit measurement.
7. Particular attention needs to be paid to increasing doubts as to the usefulness of measuring, exclusively or even primarily, at the “reaction” level. Data from the reaction level are often misleading and responses may have little relationship with future learning – the limitations of the traditional “smile” sheets and questionnaires should be borne in mind, whatever method is applied. UN system agencies would surely benefit from greater emphasis on the “learning level”, i.e. knowledge acquisition, and questionnaires should be framed with this in mind. Evaluating the levels of “behaviour” (worksite implementation) and “results” (impact on organization) – to use Kirkpatrick’s formula – requires resources which may prove beyond the capacity of public international organizations.

8. The new evaluation “culture” (Duignan et al.) encourages the measurement of outcomes (end results) rather than only outputs (activities done) and this should be our ambition as well, despite the resource limitations noted above. As far as feasible our aim should be to measure the consequences of our outputs. However, for the measurement of outputs attention must also be devoted to creating adequate indicators. This is in itself a major hurdle for evaluators.
9. Attention should also be focused on the wide range of organizational measures, outside of whatever methodological approach may be preferred, which could support an evaluation framework; some of these have been suggested in the sections above on “requirements” and “findings”.
10. As difficult as the task is one should perhaps recall one of Murphy’s Laws: “Left to themselves, things tend to go from bad to worse”, even though one is humbled by another Murphy Law which reminds us that “when all is said and done, a lot more is said than done.”

## V. Annex on UNEG’s document on Evaluation in the UN System

### Evaluation in the UN System<sup>55</sup>

#### **a. The Purpose of Evaluation**

Evaluation is an instrument that predominates in the national and international public sectors as a means to ensure substantive (rather than financial) accountability of the investments made, and as a basis for learning to improve the relevance and quality of future actions.

Within the specific context of the UN, evaluation helps to ensure the accountability of the various UN bodies, their managers and staff, to the General Assembly (GA) and/or their respective Governing Bodies, as well as to national stakeholders (particularly national governments). At the same time, it supports reflection and learning by the Member States, Governing Bodies, management and staff, as well as national stakeholders, on the relevance, effectiveness, efficiency, impact and sustainability of UN activities, so as to be able to improve on them.

Evaluation serves this dual purpose through the provision of reliable and credible evaluative evidence, analyses and information to Member States, the Secretary-General, programme managers, staff, and national stakeholders, on the activities of the UN system and their impact. These evaluation outputs are provided in the form of evaluation reports, briefings, various information exchanges and other evaluation products; including the act of conducting or participating in the evaluation itself. In order to be of use, they have to be

---

<sup>55</sup> This document is an extract from the paper produced by the UNEG Working Group on Evaluation and Oversight. The full paper was presented at the UNEG Annual General Meeting, 18-20 April 2007.

provided in a timely manner, in relation to the different organizations' programme planning, budgeting, implementation and reporting cycles.

Because evaluation has to simultaneously support both accountability and learning at different levels of governance, oversight, management, and operations, the conduct of evaluation has to be carried out at these different levels within each organization. Typically, in large, complex organizations with decentralized, global operations, evaluation is divided between centralized and decentralized functions. Due to the fragmented nature of governance in the UN, with various subsidiary organs and independent boards responsible for the various funds, programmes and specialized agencies, there is no single center for the production or consumption of evaluation.

## **b. How Evaluation Works in the UN**

As can be seen in the Governance and Oversight overview<sup>56</sup>, the United Nations system consists of various entities with diverse mandates and governing structures that aim to engender principles such as global governance, consensus building, peace and security, justice and international law, non-discrimination and gender equity, sustained socio-economic development, sustainable development, fair trade, humanitarian action and crime prevention.

The heterogeneity of mandates of the UN System organizations, covering normative, analytical and operational activities, combined with the requirement for evaluation to be carried out at different levels within each organization, has resulted in a diverse set of arrangements for the management, coordination and/or conduct of evaluation in the UN. In some cases, there is a dedicated evaluation entity established, in other cases, the evaluation entity is established within the organization's oversight entity. Others have established the evaluation entity within a programme management, policy, strategic planning or budgeting entity, and yet others have established it within an entity dedicated to research, learning, communications or other operational functions. A few have yet to establish any kind of evaluation capacity, but their management or operational staff may nevertheless, be involved in the conduct self-evaluations.

The regulations that currently govern the evaluation of United Nations activities were promulgated on 19 April 2000 in the Secretary General's bulletin (PPBME). Similar regulations and policies have been issued in recent years in several UN system organizations. For the autonomous organizations that are part of the UN system, each is governed by their own regulations and policies. In 2005, the heads of evaluation of 43 UN entities, under the auspices of the UN Evaluation Group (UNEG), adopted a common set of norms and standards for evaluation in the UN system.

### ***i. UN System-Wide Evaluation***

In support of evaluation for the UN System as a whole, there is only one entity in the UN that has a system-wide evaluation mandate from the GA; the Joint Inspection Unit.

The Joint Inspection Unit (JIU) was created on an experimental basis by the GA in 1966 and in 1976 was established as a standing subsidiary organ, responsible to the GA and the

---

<sup>56</sup> <http://www.uneval.org/index.cfm?module=Library&page=Document&DocumentID=5656>.

competent legislative organs of those specialized agencies and other international organizations within the United Nations system that have accepted its statute.

The JIU Inspectors have the broadest powers of investigation in all matters having a bearing on the efficiency of the services and the proper use of funds. Towards these ends, they may make on-the-spot inquiries and investigations. They are mandated to provide an independent review through inspection and evaluation aimed at improving management and methods and at achieving greater coordination between organizations. Its reports are addressed to the one or more organizations concerned or to all the organizations when the subject is of interest to the System as a whole, for consideration by the competent legislative organs of the organizations concerned. Notes and confidential letters are submitted to executive heads for their own discretionary use. The JIU produces about nine reports a year (ranging from 6 to 15), and these are a mix of system-wide, thematic and agency specific topics. Given its limited size, scope of work and ad hoc approach to topic selection, though a valuable source of independent analysis and information for the GA, its current programme does not constitute an adequate system-wide evaluation function.

In addition to the JIU, the UN has other ad hoc, as well as standing, arrangements for reviews, which might constitute a form of evaluation, of specific thematic or cross-agency issues system-wide. By way of example, these include –

- Ad hoc GA mandated or SG initiated reviews that are carried out by the SG or specific UN departments; for example, the High Level Panels established to evaluate - Governance and Oversight in the UN, UN System-wide coherence, UN Reform, Peace Operations, Threats, Challenges and Change, and so on. These reviews are assigned to the appropriate lead UN agency; e.g. DM, CEB, EOSG, DPKO, DPA, etc.
- The Triennial Comprehensive Policy Review, which evaluates the effectiveness and efficiency of the United Nations development system's assistance to national efforts of developing countries, specifically in their efforts to pursue their priorities and meet their needs in the context of the Millennium Declaration and other global conferences and summits. This review is assigned to the Department for Economic and Social Affairs (DESA) and is submitted to the GA / ECOSOC for consideration.

## *ii. Evaluation in the UN Secretariat*

With regard to evaluation of UN Secretariat programmes, the rules and regulations of the PPBME apply for all of them. Given the heterogeneity and size of Secretariat programme activities, in order to ensure that all programme activities are evaluated, evaluation is decentralized to the programme level, and each Secretariat programme is required to conduct regular, periodic evaluation of all activities. Moreover, in some cases, individual programmes are mandated by intergovernmental bodies that directly oversee the substantive and operational aspects of the respective programmes to conduct specific evaluations and report to them, so as to ensure effective and substantive intergovernmental oversight at the programme level. Individual programmes also receive requests from donor agencies to conduct evaluations of activities supported by voluntary funding provided by those donor agencies.

To respond to the above needs, currently, 15 of the 28 Secretariat programmes have a dedicated office, team or unit supporting evaluations; these are – UNCTAD, UNEP, UNHABITAT, UNODC, ECA, ECE, ESCAP, ECLAC, ESCWA, UNCHR, UNHCR, OCHA, DPI, OIOS and UNON. Some of these have, in addition to the PPBME, established programme specific evaluation policies (OCHA, UNEP, UNODC, ESCAP and UNHCR). The remaining Secretariat programmes do not have any dedicated evaluation office, team or unit, though self-assessment activities are conducted by management or programme staff for the purpose of mandatory reporting. In a few cases, there has been little or no evaluation at all, of programme activities.

The central evaluation function of the Secretariat is assigned to OIOS, where evaluation complements its other oversight functions of investigation, audit and inspections, by focusing on broad issues of relevance, effectiveness, efficiency and impact of Secretariat programmes and activities. The Evaluation Section of OIOS is mandated by the GA to conduct in-depth and thematic evaluations on the work of the Secretariat programmes, as well as to establish guidelines for the conduct of self-evaluation by the programmes, and to provide methodological support. OIOS evaluations are considered by the CPC, and as appropriate, by the Main Committees of the GA. Once endorsed by the GA, OIOS recommendations are mandatory and subject to triennial reviews for compliance.

### ***iii. Evaluation in the UN Funds and Programmes***

The UN Funds and Programmes that are not part of the Secretariat; i.e. UNDP, UNFPA, UNICEF and WFP, have well established evaluation offices and evaluation policy. Again, given the size of these programmes and their global reach, with their complex activities and stratified management arrangements, it has also been necessary to establish decentralized and centralized evaluation functions within each of them. Typically, the central evaluation offices conduct programme, thematic and country-level evaluations, while providing methodological guidance to decentralized evaluation functions at the country and project levels, which are carried out by programme managers, country directors, and project managers. The central evaluation offices of these UN funds and programmes report to their respective Governing Bodies (or Executive Boards) through (or simultaneously to) their respective Chief Executives.

### ***iv. Evaluation in the UN Specialized Agencies and Funds***

Similarly, most of the autonomous specialized agencies have well established evaluation offices, each with their own evaluation policy. Some of these – IFAD, IBRD, IMF, IFC and GEF, have established central evaluation offices that have a high degree of independence; i.e. the evaluation offices report directly to the Executive Boards, have heads of evaluation appointed by the Board, and have budgets approved independently by the Board. Several – ILO, FAO, UNESCO, WIPO, WHO, WMO and UNIDO, have established central evaluation offices or units with operational independence and clear evaluation policy in line with UNEG norms and standards, and reporting to the Heads of the Organization, if not directly to the Governing Bodies. There are a few without central evaluation capacity (ICAO, UPU, and IAEA) but that have evaluation capacity decentralized, or embedded, within management or operational structures. Finally there are a few that do not seem to have any established evaluation capacity – ITU, IMO, OPCW, UNFIP and UNWTO (Tourism) and WTO (Trade).

*v. Evaluation in UN Research and Training Institutes*

The following are UN research and training institutes - UNDIR, UNICRI, UNRISD, INSTRAW, UNITAR, UNSSC and UNU. No information available with regard to evaluation in these institutions.

***vi. Placement of Evaluation Functions in United Nations and related organizations (56)***

Placement	Organizations
<p>Entities with Independent Stand-Alone Evaluation Units (9) (16.7%); i.e. the Head of Evaluation reports directly to either the Head of Organization or the Governing Body, and the Evaluation Unit is located separate from policy, management and operational units.</p>	<ol style="list-style-type: none"> <li>1. GEF - Evaluation Office</li> <li>2. IFAD - Office of Evaluation</li> <li>3. IMF - Independent Evaluation Office</li> <li>4. WFP - Office of Evaluation</li> <li>5. World Bank (IBRD &amp; IFC) - Independent Evaluation Group</li> <li>6. UNDP - Evaluation Office</li> <li>7. UNCDF (associated fund of UNDP) - Evaluation Unit</li> <li>8. UNV (associated prog. of UNDP) - Evaluation Unit</li> <li>9. UNAIDS - Evaluation Department</li> </ol>
<p>Entities with Evaluation co-located with Oversight Units (9) (16.7%)</p>	<ol style="list-style-type: none"> <li>1. OIOS - Evaluation Section/MECD</li> <li>2. UNEP - Evaluation and Oversight Unit</li> <li>3. UNFPA - Division for Oversight Services (DOS)</li> <li>4. UNESCO - Evaluation Section of the Internal Oversight Service (IOS)</li> <li>5. WIPO - Evaluation functions within the Internal Audit and Oversight Division (IAOD)</li> <li>6. ICAO - Office for Programmes Evaluation, Audit and Management Review</li> <li>7. IMO – Member State Audit and Internal Oversight Unit (Includes Evaluation?)</li> <li>8. WHO - Evaluation and Performance Audit Office of Internal Oversight Services</li> <li>9. WMO – Evaluation and Performance Audit Service, Internal Oversight Office</li> </ol>
<p>Entities with Evaluation co-located with Programme Policy, Management, Planning and/or Monitoring Units (17) (30.4%)</p>	<ol style="list-style-type: none"> <li>1. UNCTAD – Programme Planning and Assessment Unit (PPAU) and ITC Evaluation Unit</li> <li>2. UNODC - Independent Evaluation Unit in the Division for Policy Analysis and Public Affairs</li> <li>3. ECA - Programme Monitoring and Evaluation Section (PMES)</li> <li>4. ESCAP - Programme Planning, Budget and Evaluation Section</li> <li>5. ECE - Programme and Evaluation?</li> <li>6. ECLAC - Programme Planning and Evaluation Unit</li> <li>7. ESCWA - Programme Planning and Technical Cooperation Division</li> <li>8. UNHCR - Evaluation and Policy Analysis Unit</li> <li>9. UN-HABITAT - Monitoring and Evaluation Unit?</li> <li>10. OCHA - Evaluation and Studies Unit in the Policy Development and Studies Branch</li> <li>11. UNCHR – Project Management and Technical Cooperation Unit</li> <li>12. UNICEF - Evaluation Office, Programme and Strategic Planning</li> <li>13. FAO - Evaluation Service in the Office of Programme, Budget and Evaluation</li> <li>14. ILO - Evaluation Unit in Mgt and Admin Sector</li> <li>15. UNIDO - Evaluation Group under the Bureau for Organizational Strategy and Learning</li> <li>16. IAEA - Office of Programme Support and Evaluation</li> <li>17. CTBTO - Evaluation Section, Office of Executive Secretary</li> </ol>
<p>Entities with Evaluation co-located with Research and/or Learning Units (3) (5.4%)</p>	<ol style="list-style-type: none"> <li>1. DPI - Evaluation and Communications Research Unit</li> <li>2. UNIFEM (associated fund of UNDP) - Learning Unit</li> <li>3. DPKO – Peacekeeping Best Practices Unit</li> </ol>

<p>Entities with no dedicated evaluation units (18) (32.1%)</p>	<ol style="list-style-type: none"> <li>1. EOSG (UNOG, UNOV, UNLOAA, and UNON – UNON has a Compact Team that has evaluation?)</li> <li>2. DGACM – Proposed Monitoring and Evaluation Unit</li> <li>3. DPA</li> <li>4. DDA</li> <li>5. OOSA</li> <li>6. OLA</li> <li>7. DESA</li> <li>8. OHRLLS</li> <li>9. OSAA (NEPAD)</li> <li>10. UNRWA – Proposed Evaluation Unit under Director of Operations</li> <li>11. DM</li> <li>12. DSS</li> <li>13. ITU</li> <li>14. UPU</li> <li>15. UNWTO (Tourism)</li> <li>16. OPCW</li> <li>17. UNFIP</li> <li>18. WTO</li> </ol>
---	--

### **c. How Evaluation is Coordinated in the UN System**

Currently, there is no centralized coordination and/or planning of evaluation for the UN system as a whole. However, to some extent, “sub” system planning and coordination is being done by – OIOS for the Secretariat programmes, UNDP for the development related funds and programmes, and OCHA for humanitarian-related activities. There does exist an inter-agency working group (called the United Nations Evaluation Group (UNEG). Established in January 1984 (originally under the name of the Inter-Agency Working Group on Evaluation), UNEG’s objective is to provide a forum for the discussion of evaluation issues within the UN System and to promote simplification and harmonization of evaluation reporting practices among UNDP and the executing agencies. To date, UNEG has 43 member organizations. UNDP chairs UNEG and provides the Secretariat facilities. In 2005, UNEG adopted a common, UN system-wide set of norms and standards for evaluation. These norms and standards are currently being adapted and used by UN system organizations as appropriate for their particular evaluation needs.

#### **Notes by consultant:**

UNEG Norms & Standards for Evaluation, April 2005, are downloadable in 5 official UN languages at [www.undg.org/index.cfm?P=259](http://www.undg.org/index.cfm?P=259)

The World Bank lists the “Existing Evaluation Principles and Standards”, which include those set forth by OECD/DAC, multilateral development banks, the AEA, etc. at <http://go.worldbank.org/XCS88M5QZ0>.