

**EVALUATION
REPORT**

**THE QUALITY OF EVALUATIONS
SUPPORTED BY
UNICEF COUNTRY OFFICES
2000-2001**

**EVALUATION OFFICE
SEPTEMBER 2004**

For every child
Health, Education, Equality, Protection
ADVANCE HUMANITY



**EVALUATION
REPORT**

**THE QUALITY OF EVALUATIONS
SUPPORTED BY
UNICEF COUNTRY OFFICES
2000-2001**

The Quality of Evaluations Supported by UNICEF Country Offices, 2000-2001
© United Nations Children's Fund (UNICEF), New York, 2004

UNICEF
3 UN Plaza, NY, NY 10017
September 2004

This evaluation was undertaken by Rideau Strategy Consultants Ltd., Ottawa, Canada. UNICEF Evaluation Office commissioned this evaluation which was completed in June 2003, with the exception of Appendix 6. Dr. Kenneth Watson led the evaluation and he was assisted in this evaluation by Dr. Anne Perkins, Ms. Julia Paton, Ms. Zahra Boodhwani and Mr. André Bernier. The issue of this report was delayed to allow the addition of Appendix 6 which details UNICEF's responses to recommendations. The contents of this report do not necessarily reflect the policies or the views of UNICEF.

The text has not been edited to official publication standards and UNICEF accepts no responsibility for errors.

The designations in this publication do not imply an opinion on legal status of any country or territory, or of its authorities, or the delimitation of frontiers.

For further information, please contact:

UNICEF, 3 United Nations Plaza, H-2F
New York, NY 10017, USA
Tel. (1 212) 824-6467
Fax (1 212) 824-6492

TABLE OF CONTENTS

Executive Summary	i
Résumé Analytique	ix
Resumen Ejecutivo	xvii
1. Introduction	1
1.1. Objectives	1
1.2. Methodology	2
1.3. Scope and Limitations	4
1.4. Research Ethics	4
1.5. Previous Reviews of Evaluation Quality	4
2. UNICEF Evaluation Activity	7
2.1. Country Office Evaluation Activity	7
2.2. Accessibility and Use of Evaluations	8
3. Quality of Evaluation Design	9
3.1. Types of Evaluations	9
3.2. Quality of Terms of Reference	11
3.3. Evaluation Methodologies	12
4. Quality of Evaluation Management	15
4.1. Purpose and Timing	15
4.2. Research Ethics	15
4.3. Quality Control of Evaluations	18
4.4. Capacity Building	21
4.5. Participation of Children and Women in Evaluation	22
5. Coverage of Issues	25
5.1. Rights and Equity	25
5.2. Usefulness	26
5.3. Context and Attribution	27
5.4. Analysis of Costs	27
5.5. Analysis of Outputs	28
5.6. Analysis of Outcomes	29
5.7. Analysis of Sustainability	30
5.8. Analysis of Sustainability	30
6. Reporting and Action	33
6.1. Completeness, Clarity, and Accuracy	33
6.2. Quality of Findings, Lessons, and Recommendations	34
6.3. Link to Action	36
7. Findings	37
7.1. Overall Quality	37
7.2. Priorities and Next Steps	37
7.3. List of Recommendations	40
APPENDICES	45
Appendix 1: Evaluation Report Quality	47
Appendix 2: Country Office Feedback on Evaluation Quality	57
Appendix 3: Model Table of Contents for an Evaluation of Strategy	85
Appendix 4: Model Table of Contents for a Performance Evaluation	89
Appendix 5: Terms of Reference	91
Appendix 6: UNICEF Response to specific recommendations	97

EXECUTIVE SUMMARY

The Quality of Evaluations Supported by UNICEF Country Offices, 2000-2001

Purpose

This study evaluates the quality of UNICEF evaluations conducted and commissioned by its country offices in 2000-2001. It is UNICEF's second review of evaluation quality this decade. It also follows various thematic reviews, including a review of UNICEF-supported education evaluations.¹

The objectives of this study are to assess the quality of evaluations supported by UNICEF's country offices, to see whether progress has been made since the last review, and to recommend how quality might be improved.

Context

UNICEF has a decentralised accountability structure for evaluation.² The UNICEF country offices undertake evaluations of programmes and projects for oversight and learning. The resulting reports are the subject of this study. Although the UNICEF Regions and Evaluation Office at New York headquarters have important roles, their work is outside the scope of this study.

Criteria of Quality

Thirty criteria of evaluation quality were applied in this study. These criteria cover evaluation quality at all stages of the evaluation project cycle, from design to use. They are listed in Appendix 1, along with a statistical summary of the results of the quality review in each case. The criteria were identified from the published evaluation standards of professional evaluation associations, from literature³, especially previous reviews of evaluation quality by UNICEF⁴, USAID⁵, and CIDA⁶, and from discussion among the professional evaluators in UNICEF and on the consulting team.

A questionnaire containing 23 questions, some structured and some open-ended, was used to gather information from UNICEF country offices (Appendix 2).

In addition, a survey of staff in a 50% sample of UNICEF country offices [47] was undertaken. The response was close to 100%.

¹ Chapman, D. Desk review of 185 evaluations.

² UNICEF "Report on the Evaluation Function in the Context of the MTSP", E/ECE/200210.

³ For example, see Thomas Black, *Evaluating Social Science Research: An Introduction*, London, Sage, 1993.

⁴ C.G. Victora, et al. *A Systematic Review of UNICEF-Supported Evaluations and Studies, 1992-1993*, UNICEF, New York, Working Paper No. 3, June 1995.

⁵ Center for Development Information and Evaluation, U.S. Agency for International Development, *Review of the Quality of A.I.D. Evaluations: FY 1987 and FY 1988*, Occasional Paper No. 19, USAID, Washington, D.C., May 1989. Also, *Action Plan: Upgrading Central Evaluation in A.I.D.*, 1991.

⁶ Canadian International Development Agency, *A Study of the Quality of CIDA's Bilateral Evaluations: 1988-89 and 1989-90*, Ottawa, June 1991 (per Dr. Kenneth Watson).

Scope, Limitations, and Methodology

A 50% random sample of the UNICEF evaluation reports that had been submitted to New York Headquarters by the country offices in the years 2000 and 2001 was reviewed. Of those 97 reports, 75 were found to be evaluations, strictly defined. For approximately one third of the reports, we reviewed the Terms of Reference for quality as well.

Finally, the assessment team also examined the quality of 31 Terms of Reference for our sample of evaluations — about one third — all that were available at New York headquarters for this set of studies.

The quality assessors were independent. The review was commissioned by competitive tender, and no reviewer had been involved with any of the evaluations. The research design for this study was not subject to ethical review. However, since the work was done entirely from existing UNICEF reports, and from information proffered by UNICEF staff in the course of their professional duties, no issues of informed and competent consent were expected to arise. The only ethical issue was protection of informant confidentiality during this work.

This study had two significant limitations. First, it was based solely on documents, interviews at New York headquarters, and a written questionnaire for country offices. Country offices were not visited, and country office staff not interviewed in depth. Evaluation procedures were not observed directly, nor files reviewed. Partners and users of UNICEF evaluations were not interviewed. Our judgements were based solely on the quality of the evaluation reports and their written Terms of Reference. Consequently we had a narrow base for our recommendations about how evaluation quality might be improved, and any recommendations on how better management of evaluation might improve quality should be read in this light.

The second important limitation of this study was that the evaluation reports examined were sampled from those submitted to New York headquarters by the country offices, and not all evaluations supported by country offices were submitted. We expect that poor reports are less likely to be submitted to headquarters. Respondents to the questionnaire indicated that their offices had completed approximately 75 evaluations during 2000 or 2001 that had not been submitted to New York headquarters. However, there may have been more. On the other hand, some reports not submitted may not have been evaluations by our definition. The reader should keep in mind that selective submission of evaluation reports by country offices may have resulted in bias in our sample.

Findings

Overview

It was found that UNICEF evaluations are not consistent in quality. About one in five are excellent, but the worst third are sufficiently poor to constitute a serious problem.

Best and Worst Aspects

The five aspects of quality on which the UNICEF evaluations did best were:

- The objectives of many evaluations, and the questions to be answered, were often stated fully and clearly.
- Many evaluation reports were clear, transparent, and easily accessible to the reader. The best were concise, well-organised, and logical, with clearly written text, supported by tables, figures, and descriptive headings, and led by an executive summary.

- The objectives of the evaluation, and the questions to be answered, often reflected UNICEF’s mission and approach to programming, including protection of children’s rights, promotion of their welfare, and gender equality (gender being the weakest of these).
- Recommendations were often well based on evidence and analysis.
- The qualitative and quantitative information gathered by many evaluations was, in aggregate, adequate to answer the evaluation questions.

The five criteria of quality on which the UNICEF evaluations did worst were:

- Costs were not well described, and were seldom compared with results.
- The “outputs” of the programme or project were often not adequately described or measured and, with this missing link, the causal chain from activities to outcomes was broken.
- Ethics review was seldom undertaken at the research design stage, and the topic of research ethics was seldom addressed in the reports. It is, of course, vital that the evaluation design be ethical and include ethical safeguards where appropriate, including protection of the dignity, rights, and welfare of human subjects, particularly children, and respect for the values of the beneficiary community. We have no opinion on whether there were any ethical problems with the research, such as competence or informed consent, but simply note that the evaluations seldom addressed the topic. The evaluators seldom made a statement about how their objectivity and independence were ensured.
- The evaluations were generally parochial. The degree to which the project, programme, or initiative might be replicable in other contexts often was not described.
- Lessons learned often were not generalised beyond the immediate intervention being evaluated to indicate what wider relevance to UNICEF there might be.

Priorities

To improve the quality of its evaluations, UNICEF could focus on risk or excellence, or, of course, both, if sufficient resources can be mobilised.

Option 1: Maximise UNICEF’s influence by focusing evaluation efforts on producing a relatively small number of excellent evaluations of intervention strategies in vital areas of intervention, and with wide replicability.

Option 2: Minimise risk by upgrading the poorest third of UNICEF evaluations to minimum professional standards.

The tools to upgrade the worst third of evaluations should be appropriately simple. They might include generic frameworks for evaluation terms of reference (including guidelines for processes, timeframes, and budgets), and standard Tables of Contents for the two main types of evaluations [performance evaluations and evaluations of intervention strategy].

At the other end of the quality spectrum, the problem and the appropriate response are quite different. UNICEF has produced many evaluations that are good but not excellent. Achieving excellence, starting from this base of good work, would be more difficult and expensive than the “minimum standard” option described above. Three things would be important:

- Engaging the best evaluation professionals, which would probably be considerably more expensive than some country offices think they can afford.
- Insisting on rigorous research designs that are much less impressionistic than UNICEF evaluations often have been.
- Developing methodologies for evaluations of various strategies for rights-based interventions. The cultural, political, and legal dimensions of rights-based interventions make their evaluation particularly challenging, and UNICEF often would be breaking new ground in evaluation methodology.

Achieving excellence would require highly trained evaluation managers, as well as highly qualified consultants, and larger budgets to enable more thorough primary data collection and analysis. All this raises issues of how much UNICEF is able and willing to pay for evaluations. These issues are beyond the scope of this study. Perhaps the only way to achieve excellence within an affordable budget is by doing a limited number of broadly relevant evaluations of intervention strategies in key topic areas, each year, each involving several country offices. The design and coordination of such evaluations might require leadership from UNICEF regions and headquarters.

Another approach to excellence, within the constraints of economy, is to undertake more evaluations of intervention strategies jointly with other development agencies, including UNDAF partners, the multilateral development banks, and the major bilateral agencies. Cases where this was done for performance evaluations of jointly funded interventions tended to be better quality than UNICEF-alone efforts. Since many areas of rights are of common interest, it might be possible to institute a series of joint evaluations of intervention strategies in key areas with one or more partners among the international agencies.

National Partner Capacity Enhancement

The suggestions above relate to UNICEF’s own evaluation capability and performance. They will not improve the evaluation capability of host-country agencies. Attempts to upgrading local capability by involving a few local staff and/or local consultants in UNICEF evaluations in minor roles may not be very effective. Most developing countries need more structural assistance to improve their evaluation capabilities, such as forming a national evaluation association with affiliation with an international professional association. Again, other international agencies are interested and involved in enhancing the evaluation capacities of national partners, and UNICEF should undertake joint capacity-enhancement projects wherever possible.

Main Recommendations

Our general recommendation is, of course, that everything possible be done to improve the quality of UNICEF's evaluations. However, we are aware that this study is only partly adequate as a basis for recommendations on how to do this. It does not cover all UNICEF evaluations, nor does it examine UNICEF's evaluation resources, systems, and practices. That said, we make the following main recommendations:

1. UNICEF Evaluation Office should formulate an action plan for evaluation quality improvement in response to this study.

This study should be complemented by a study of the resources and organisation of UNICEF evaluation. Given the persistence of the same quality problems over a long period of time, some systemic changes might be in order. Pending the outcome of such a complementary study, consideration should be given to balancing UNICEF's decentralised country office-focused evaluation system with strengthened requirements for review of evaluation research designs outside the initiating country office at the time the terms of reference are being formulated. Better research designs are probably the single thing most likely to improve the quality of UNICEF evaluations. Each evaluation study should have a methodological review and challenge by a peer outside the country office. For each evaluation budgeted at over \$25,000, this review should be based on a full evaluation framework.⁷

2. Since virtually all UNICEF evaluations involve human subjects, the country office evaluations should be subject to stronger requirements for ethics review before implementation. We suggest that UNICEF Evaluation Office state an evaluation research ethics review policy, and that an appropriate system of evaluation research ethics review be established. This would include a policy on ethics review of Terms of Reference, subject competence and informed consent, and a policy on adverse event reporting. We do not believe that it is sufficient for the Evaluation Office to rely on other UNICEF policy statements, in this respect, or that the Technical Note⁸ extant is sufficient. In addition, each evaluation report should contain a statement of how the objectivity and independence of the evaluators was ensured.
3. Since country office evaluation reports continue to exhibit many of the deficiencies found in 1995 regarding evaluation reports in the early 1990s, despite an improved evaluation policy, guidelines and technical notes, we recommend that certain things be made mandatory content for every evaluation study (this requirement reinforced by model Tables of Contents for evaluation reports). Evaluation reports should follow a standard format, unless there is a good reason for varying it.⁹ UNICEF terms of reference for evaluation studies should include a draft Table of Contents for the evaluation report,

⁷ An "evaluation framework" comprises a concise profile of the project or programme to be evaluated, including a logic model that underpins its rationale (a logical framework analysis or another appropriate model), a list of the questions to be answered by the evaluation team, and a description of methodology that is appropriate to the type of study [strategy evaluation or performance evaluation) and to the questions.

⁸ UNICEF Evaluation Office, Technical Note Number 1, "Children Participating in Research, Monitoring and Evaluation – Ethics and Your Responsibilities as a Manager", April 2002.

⁹ Standardization of the content/format of the main evaluation report would not preclude producing outputs in other formats for various dissemination pieces.

such as those models shown in Appendices 3 and 4. Standard content should include the following:

- There should be a clear description of what is being evaluated, including sufficient background and context to enable a reader unfamiliar with the country and programme to fully understand it, and explicitly describing the relevance and replicability in other contexts of the project/programme and its evaluation. All UNICEF evaluations of intervention strategy should address the wider relevance of lessons learned, and the replicability and scalability of the successful aspects of the project or programme. There should be a full profile of the intervention and its context, an attribution analysis where appropriate, and a consideration of replicability, scalability, sustainability, and environmental aspects. There should be a clear statement whether the study is a situation analysis, an evaluation of alternative intervention strategies, and/or an evaluation of UNICEF's or its partner's past performance in a programme or project.
 - An analysis of costs, and, where appropriate, of efficiency, should be part of all evaluations. UNICEF should have guidelines on how to estimate the full cost of a project or programme, including UNICEF staff time, contracts, costs of partners, and costs of participants.
 - There should be detailed measurement of the programme/project outputs. To enable this, UNICEF contracts/agreements with implementing agents should be output-performance-based. This base of clear output-performance agreements is essential to a results-based approach to good management, and to enabling good performance evaluation.
 - If the project or programme is expected to be sustained, with or without continued UNICEF funding, then the mechanisms for ensuring this should be explained. If commitments from others are necessary, then these should be described. Self-sustainability through cost-recovery should always be one sustainability alternative assessed.
4. The UNICEF country office should record its decision on each evaluation recommendation. The evaluation manager (focal point) should prepare an action plan for approval by the country representative. The approved action plan should become an appendix to the evaluation report before it is submitted to NY headquarters for archiving. Where an evaluation recommendation requires action by another agency, UNICEF should ask that agency for a response. To facilitate action plans following evaluation reports, draft findings, lessons, and recommendations should be subjected to challenge in an "exit workshop" that involves all major stakeholders.

In addition to these main recommendations, we make the following suggestions:

General Format and Content

1. Evaluation reports should be limited to a more –or less standard length of presentation, say 50 pages + executive summary + appendices.

2. Evaluation reports should discuss data quality, and explain the effects of reach constraints, attrition, and/or non-response.
3. An executive summary of the evaluation report should be translated into the local language[s]. This should be part of the Terms of Reference and appropriately budgeted.

General Procedures

We suggest that NYHQ Evaluation Office and Regional Offices develop a three-year rolling evaluation plan in conjunction with the country offices in order to coordinate evaluation activity and ensure adequate coverage of key issues. The plan should be updated annually.

UNICEF needs some means of frequently updating country office awareness of what evaluations are being started and completed in other country offices.

We suggest that the full text of completed evaluations from the current year and two previous years be accessible on the Internet in Adobe format, and key worded for easy search and access.¹⁰ We believe that, to facilitate openness and to give an incentive to improve quality, all UNICEF evaluations that cost more than, say, \$10,000 should be available to the public on the open website.

1. UNICEF should require that an electronic copy of all Terms of Reference should be submitted to the Regional Office NYHQ before an evaluation is contracted and an electronic copy of the evaluation report afterwards.
2. To assure minimum quality standards are met, UNICEF should consider instituting more requirements for management, evaluation, and sector expert sign-offs, especially at the evaluation research design stage. Sign-offs should be acknowledged in the final report.

¹⁰ Canadian International Development Agency; op. cit., p. 48. "Evaluation abstracts, as useful as they are in alerting a broad cross-section of staff to the existence of an evaluation report, cannot fill the understanding gap. Evaluations are complex and they demand an in-depth reading if the project or program is to be understood and the lessons grasped."

RÉSUMÉ ANALYTIQUE

La qualité des évaluations financées par les bureaux nationaux de l'UNICEF, 2000-2001

Objet

Cette étude évalue la qualité des évaluations de l'UNICEF commanditées et conduites par ses bureaux nationaux en 2000-2001. C'est la seconde étude de qualité des évaluations de l'UNICEF cette décennie. Elle intervient également à la suite d'examens thématiques variés, dont un examen des évaluations sur l'éducation financé par l'UNICEF.¹¹

Les objectifs de cette étude sont d'évaluer la qualité des évaluations financées par les bureaux nationaux de l'UNICEF pour voir si des progrès ont été accomplis depuis la dernière étude, et de recommander des façons d'en améliorer la qualité.

Contexte

L'UNICEF a pour l'évaluation une structure de responsabilités décentralisée.¹² Les bureaux nationaux de l'UNICEF entreprennent des évaluations de programmes et de projets pour la supervision et l'apprentissage. Les rapports qui en découlent sont le sujet de la présente étude. Bien que les régions de l'UNICEF et le bureau des évaluations au Siège de New York aient un rôle important à jouer, leurs travaux se situent en dehors du champ d'application de cette étude.

Critères de qualité

Cette étude a mis en application trente critères de qualité des évaluations à tous les stades du cycle du projet d'évaluation, de la conception à l'utilisation. On en trouvera la liste dans l'appendice 1, avec dans chaque cas un sommaire statistique des résultats de l'étude de qualité. Ces critères ont été identifiés à partir : 1) des normes d'évaluation publiées par les associations d'évaluateurs professionnels, 2) de la littérature publiée à ce sujet¹³, et en particulier des études semblables précédemment menées par l'UNICEF¹⁴, USAID¹⁵ et l'ACDI¹⁶, 3) des discussions menées entre évaluateurs professionnels, à l'UNICEF et dans l'équipe de consultants.

Un questionnaire de 23 questions, certaines structurées et d'autres ouvertes, a été utilisé pour recueillir des informations auprès des bureaux nationaux de l'UNICEF (Appendice 2).

¹¹ Chapman, D.: examen des dossiers de 185 évaluations.

¹² UNICEF : « Rapport sur la fonction d'évaluation dans le contexte du Plan stratégique à moyen terme », E/ECE/200210.

¹³ Par exemple, voir Thomas Black : « Evaluating Social Science Research: An Introduction », Londres, Sage, 1993.

¹⁴ C.G. Victora, et consorts : « A Systematic Review of UNICEF-Supported Evaluations and Studies, 1992-1993 », UNICEF, New York, Document de travail No. 3, Juin 1995.

¹⁵ Center for Development Information and Evaluation, U.S. Agency for International Development, Review of the Quality of A.I.D. Evaluations: FY 1987 and FY 1988, Occasional Paper No. 19, USAID, Washington, D.C., Mai 1989. Aussi, Action Plan: Upgrading Central Evaluation in A.I.D., 1991.

¹⁶ Agence canadienne de développement international, Une étude de la qualité des évaluations bilatérales de l'ACDI : 1988-89 et 1989-90, Ottawa, Juin 1991 (par le Dr. Kenneth Watson).

Champ d'application, limitations et méthodologie

Il a été procédé à un examen d'un échantillonnage pris au hasard de 50% des rapports qui ont été soumis au Siège de New York par les bureaux de pays dans les années 2000 et 2001. De ces 97 rapports, on a trouvé que 75 étaient des évaluations au sens strict du terme. Pour le tiers de ces rapports environ, nous avons également examiné la qualité du cahier des charges.

De plus, une enquête a été menée auprès du personnel de 50% des bureaux nationaux (soit 47). La réponse a été de presque 100%.

Enfin, l'équipe d'étude a aussi examiné la qualité de 31 cahiers des charges pour notre échantillonnage d'évaluations (environ un tiers), soit ce qui était disponible au Siège de New York pour cette série d'études.

Les évaluateurs de la qualité étaient indépendants. L'étude a été attribuée sur appel d'offres, et aucun des analystes n'avait été lié aux évaluations en question. Le concept de recherche pour cette étude n'était pas soumis à un examen d'éthique. Et pourtant, puisque les travaux ont été entièrement effectués à partir de rapports existants de l'UNICEF et d'informations présentées par le personnel de l'UNICEF dans l'exercice de ses responsabilités professionnelles, on n'attendait pas que se posent des problèmes de consentement éclairé et de compétence. La seule question relevant de l'éthique était la protection de la confidentialité de l'informateur pendant ces travaux.

Cette étude avait deux limitations importantes. D'abord, elle ne se fondait que sur des documents, des entretiens au Siège de New York, et un questionnaire écrit envoyé aux bureaux de pays. Les bureaux de pays n'ont pas été visités, et leur personnel n'a pas été interrogé en profondeur. Les procédures de l'évaluation n'ont pas été observées directement, les dossiers n'ont pas été examinés. Les partenaires et les utilisateurs des évaluations de l'UNICEF n'ont pas été interrogés. Nos jugements se sont fondés exclusivement sur la qualité des rapports d'évaluation et leur cahier des charges écrit. Nos recommandations sur la manière dont on pourrait améliorer la qualité de l'évaluation ont donc un fondement ténu, et toute recommandation sur la manière dont une meilleure gestion de l'évaluation pourrait en améliorer la qualité devrait être lue à la lumière de cette réalité.

La seconde limitation importante de cette étude était que les rapports d'évaluation examinés représentaient un échantillonnage de ceux qui avaient été soumis au Siège de New York par les bureaux de pays, et que tous les rapports financés par les bureaux de pays n'ont pas été soumis, car nous pensons que les mauvais rapports ont moins de chance d'être soumis au Siège. Ceux qui ont répondu au questionnaire ont indiqué que leur bureau avait effectué 75 évaluations environ en 2000 et 2001 qui n'ont pas été soumises au Siège de New York, et il peut y en avoir encore davantage. Par ailleurs, certains rapports qui n'ont pas été soumis peuvent ne pas avoir été des évaluations au sens que nous donnons à ce terme. Le lecteur devrait garder à l'esprit qu'une soumission sélective de rapports d'évaluation par les bureaux de pays peut s'être soldé par la partialité de l'échantillonnage.

Constatations

Vue d'ensemble

Nous avons découvert que les évaluations de l'UNICEF n'étaient pas d'une qualité homogène. Une sur cinq environ est excellente, mais le tiers le plus mauvais l'est suffisamment pour constituer un problème sérieux.

Les aspects les meilleurs et les pires

Les cinq aspects de qualité où les évaluations de l'UNICEF ont été les meilleures étaient les suivants :

- Les objectifs de nombreuses évaluations et les questions auxquelles une réponse devait être apportée étaient souvent complets et clairs dans leur énoncé.
- De nombreux rapports d'évaluation étaient clairs, transparents et d'accès facile pour le lecteur. Les meilleurs étaient concis, bien structurés et logiques, avec une écriture claire appuyée par des tableaux, des chiffres et des intitulés descriptifs, et introduits par un résumé analytique.
- Les objectifs de l'évaluation et les questions auxquelles il fallait apporter une réponse reflétaient souvent la mission de l'UNICEF et sa méthode de programmation : protection des droits des enfants, promotion de leur bien-être et égalité des sexes (ce dernier point étant le moins bien rendu).
- Les recommandations se fondaient souvent bien sur des preuves et des analyses.
- Les informations qualitatives et quantitatives réunies par de nombreuses évaluations convenaient dans l'ensemble pour répondre aux questions de l'évaluation.

Les cinq critères où l'UNICEF était le plus mauvais ont été les suivants :

- Les coûts n'étaient pas bien décrits et rarement comparés aux résultats.
- Souvent, les « produits » du programme ou du projet n'étaient pas convenablement décrits ou mesurés et, avec ce chaînon manquant, la chaîne de cause à effet reliant les activités aux réalisations a été rompue.
- L'examen des considérations éthiques a été rarement entrepris au stade de la conception, et la question de l'éthique de la recherche a été rarement abordée dans les rapports. Il est bien entendu vital que la conception de l'évaluation soit éthique et comporte des garde-fous éthiques lorsqu'il le faut : protection de la dignité, des droits et du bien-être des sujets humains et surtout des enfants, et respect des valeurs de la communauté bénéficiaire. Nous n'avons aucune opinion sur l'éventualité de problèmes d'éthique dans la recherche, comme la compétence ou le consentement éclairé, mais veuillez simplement noter que les évaluations n'abordaient que rarement ce sujet. Les évaluateurs ont rarement expliqué dans une déclaration la manière dont leur objectivité et leur indépendance ont été garanties.
- Les évaluations relevaient généralement d'un esprit de clocher. Le degré auquel le projet, le programme ou l'initiative pouvaient être reproduits dans d'autres contextes n'était pas décrit.
- Souvent, les enseignements tirés n'ont pas été généralisés au-delà de l'intervention immédiatement évaluée pour indiquer s'ils pouvaient avoir une pertinence plus large pour l'UNICEF.

Priorités

Pour améliorer la qualité de ses évaluations, l'UNICEF pourrait insister sur le risque ou l'excellence, ou bien sûr les deux, si des ressources suffisantes peuvent être mobilisées.

Option 1 : optimiser l'influence de l'UNICEF en orientant les efforts de l'évaluation sur la production d'un nombre relativement faible d'évaluations excellentes portant sur les stratégies utilisées dans des secteurs d'intervention cruciaux, avec une large reproductibilité.

Option 2: minimiser le risque en rehaussant le niveau du plus mauvais tiers des évaluations de l'UNICEF jusqu'à atteindre des normes professionnelles minimales.

Comme il convient ici, les outils utilisés pour rehausser le niveau du plus mauvais tiers des évaluations devraient être simples. Il pourrait s'agir de cadres génériques pour les cahiers des charges (consignes de procédure, échéanciers et budgets) et de tables des matières types pour les deux grandes catégories d'évaluations : évaluation des performances et évaluation des stratégies d'intervention.

A l'autre bout de l'éventail de qualité, le problème et la réaction correspondante sont assez différents. L'UNICEF a produit de nombreuses évaluations qui sont bonnes mais pas excellentes. Atteindre un niveau d'excellence à partir de ce bon travail serait plus difficile et plus coûteux que l'option « normes minimum » décrite ci-dessus. Trois choses seraient importantes :

- Engager les meilleurs experts de l'évaluation, ce qui serait probablement beaucoup plus cher que ce qui, pensent certains bureaux nationaux, est dans leurs moyens.
- Insister sur des schémas de recherche rigoureux qui font beaucoup moins confiance aux impressions que ce qu'on a souvent pu lire dans les évaluations de l'UNICEF.
- Elaborer des méthodologies pour les évaluations de diverses stratégies en vue d'interventions fondées sur les droits. Les dimensions culturelles, politiques et juridiques des interventions fondées sur les droits rendent leur évaluation particulièrement difficile, et l'UNICEF s'imposerait comme précurseur de nouvelles méthodologies de l'évaluation.

Atteindre l'excellence demanderait une formation exemplaire des directeurs de l'évaluation et des consultants, et des budgets plus importants pour permettre une collecte et une analyse des données primaires plus complète. Tout cela soulève le problème de ce que l'UNICEF peut et veut payer pour les évaluations. Cette question dépasse le cadre de cette étude. Peut-être le seul moyen d'atteindre l'excellence avec un budget abordable est-il de se livrer chaque année à un nombre limité d'évaluations de stratégies d'intervention qui soient pertinentes au niveau général, chacune d'entre elles faisant intervenir plusieurs bureaux nationaux. La conception et la coordination de ces évaluations pourraient demander le leadership des régions et du Siège de l'UNICEF.

Une autre démarche, dans la limite des contraintes financières, est d'entreprendre davantage d'évaluations de stratégies d'intervention en coopération avec d'autres organismes du développement : partenaires du Plan-cadre des Nations Unies pour l'aide au développement, banques multilatérales pour le développement, organismes bilatéraux importants. Dans les cas où cela a été fait pour les évaluations de performances d'interventions co-financées, le résultat a été plutôt de meilleure qualité que lorsque l'UNICEF s'est fié à ses propres efforts. Puisque de

nombreux secteurs des droits de l'homme sont d'intérêt commun, on pourrait peut-être instituer une série d'évaluations conjointes de stratégies d'intervention dans des secteurs clés, avec le partenariat d'un ou plusieurs organismes internationaux.

Amélioration de la capacité du partenaire national

Les suggestions ci-dessus se rapportent à la capacité d'évaluation et la performance propres à l'UNICEF. Elles n'amélioreront pas la capacité d'évaluation des organismes du pays d'accueil. Tenter d'améliorer les capacités locales en faisant intervenir dans des rôles mineurs quelques employés locaux de l'UNICEF ou ses consultants de l'évaluation pourrait s'avérer peu fructueux. La plupart des pays en développement ont besoin d'une aide structurelle plus élaborée pour améliorer leurs capacités d'évaluation, comme par exemple en constituant une association nationale d'évaluation affiliée à une association professionnelle internationale. Là encore, d'autres organismes internationaux s'intéressent à l'amélioration des capacités d'évaluation des partenaires nationaux et s'y emploient, et l'UNICEF devrait entreprendre des projets conjoints d'amélioration des capacités toutes les fois où c'est possible.

Recommandations principales

Notre recommandation générale est bien sûr que le maximum soit fait pour améliorer la qualité des évaluations de l'UNICEF. Toutefois, nous sommes conscients que cette étude ne convient que partiellement comme base de recommandations sur la manière de procéder. Elle ne couvre pas toutes les évaluations de l'UNICEF et n'examine pas non plus les ressources d'évaluation, les systèmes et les pratiques de l'UNICEF. Cela étant dit, nos recommandations essentielles sont les suivantes :

1. Le Bureau des évaluations de l'UNICEF devrait formuler un plan d'action pour évaluer l'amélioration de la qualité comme suite à la présente étude.

Cette étude devrait être complétée par une étude des ressources et de l'organisation de l'évaluation à l'UNICEF. Compte tenu de la persistance des mêmes problèmes de qualité sur une longue période, certains changements systémiques pourraient s'imposer. En attendant les résultats de cette étude complémentaire, on devrait songer à équilibrer le système d'évaluation de l'UNICEF qui est décentré et axé sur les bureaux de pays ; on devrait établir des conditions plus strictes d'étude des modèles de recherche pour l'évaluation, extérieurement au bureau de pays initiateur et au moment de la formulation des cahiers des charges. De meilleurs modèles de recherche sont probablement le facteur le plus à même d'améliorer la qualité des évaluations de l'UNICEF. Chaque étude sur l'évaluation devrait comporter un examen méthodologique et une critique de collègues extérieurs au bureau de pays. Pour chaque évaluation inscrite au budget à plus de 25 000 dollars, cet examen devrait se fonder sur un cadre d'évaluation complet.¹⁷

2. Puisque pratiquement toutes les évaluations de l'UNICEF mettent en jeu des sujets humains, les évaluations des bureaux nationaux devraient être assujetties à des conditions plus strictes d'éthique qui seront examinées avant la mise en œuvre. Nous suggérons que le Bureau des évaluations de l'UNICEF énonce une politique d'examen

¹⁷ Un « cadre d'évaluation » comprend une description concise du projet ou du programme à évaluer, dont un modèle logique qui étaye sa logique interne (analyse de cadre logique ou autre modèle approprié), une liste des questions auxquelles l'équipe d'évaluation doit répondre, et une description méthodologique convenant au type d'étude (évaluation des stratégies ou des performances) et aux questions.

éthique des travaux de recherche en évaluation, et que soit institué un système d'examen approprié. Ceci consisterait en une politique relative à l'examen éthique du cahier des charges, de la compétence des sujets et du consentement éclairé, et une autre sur les rapports relatifs aux événements indésirables. Nous ne croyons pas à cet égard qu'il soit suffisant que le Bureau des évaluations s'en remette à d'autres déclarations de politique de l'UNICEF, ou que la note technique existante¹⁸ à ce sujet suffise. De plus, chaque rapport d'évaluation devrait contenir une déclaration sur la manière dont l'objectivité et l'indépendance des évaluateurs a été garantie.

3. Puisque les rapports d'évaluation des bureaux de pays continuent à révéler beaucoup des lacunes constatées en 1995 sur les rapports d'évaluation au début des années 1990, nous recommandons, malgré l'amélioration de la politique d'évaluation et outre les consignes et notes techniques, que certaines choses figurent obligatoirement au contenu de chaque étude d'évaluation (cette condition étant renforcée par des tables de matières types pour les rapports d'évaluation). Les rapports d'évaluation devraient suivre une formule standard à moins qu'il y ait une bonne raison pour la modifier.¹⁹ Les cahiers des charges de l'UNICEF pour les études sur l'évaluation devraient comporter un projet de table des matières pour le rapport d'évaluation, dont des modèles sont donnés aux appendices 3 et 4. Le contenu type devrait en être le suivant :

- Il devrait y avoir une description claire de ce que l'on est en train d'évaluer, et suffisamment d'histoire et de contexte pour qu'un lecteur connaissant mal le pays et le programme puisse comprendre entièrement ; ce devrait être une description explicite de la pertinence et de la reproductibilité dans d'autres contextes du projet ou programme concernés, et de leur évaluation. Toutes les évaluations des stratégies d'intervention de l'UNICEF devraient aborder la question plus vaste de la pertinence des enseignements tirés, et celle de l'adaptabilité, dans d'autres lieux ou à une autre échelle, des aspects réussis du programme ou du projet. Il devrait y avoir une description complète de l'intervention et de son contexte, une analyse d'attribution quand cela s'impose, et une réflexion sur les questions d'adaptabilité et de durabilité ainsi que sur les aspects écologiques. Il devrait y avoir une définition claire et affirmée de l'étude : analyse de situation, évaluation des alternatives dans les stratégies d'intervention, ou une évaluation des performances passées de l'UNICEF ou de celles de son partenaire dans un programme ou un projet.
- Une analyse des coûts et, quand cela convient, de l'efficacité au niveau des coûts devrait faire partie de toutes les évaluations. L'UNICEF devrait avoir des directives sur la manière de faire une estimation du coût total d'un projet ou d'un programme, comprenant le temps du personnel, les contrats, les coûts incombant aux partenaires et aux participants.
- Il devrait y avoir une mesure détaillée des produits générés par le programme ou le projet. Pour ce faire, les contrats ou accords de l'UNICEF avec des agents de mise en œuvre devraient se fonder sur la rentabilité des produits. Cette base d'accord clairement orientée vers la rentabilité des produits est essentielle à une méthode de

¹⁸ Bureau des évaluations de l'UNICEF Evaluation Office, Note technique numéro 1, « Participation des enfants à la recherche, au suivi et à l'évaluation : les questions d'éthique et vos responsabilités de directeur », avril 2002.

¹⁹ La standardisation du contenu ou de la formule du rapport d'évaluation principal n'empêcherait pas que l'on produise des rapports selon d'autres formules pour la diffusion de divers articles.

bonne gestion axée sur les résultats ainsi qu'à une bonne évaluation des performances.

- Si l'on s'attend à ce que le projet ou le programme soient pérennisés, avec ou sans le financement continu de l'UNICEF, les mécanismes le garantissant devraient être décrits. S'il faut pour cela que d'autres parties s'engagent, il faut les décrire aussi. La viabilité grâce au recouvrement des coûts devrait toujours être une alternative à évaluer.
4. Le bureau de pays de l'UNICEF devrait enregistrer sa décision à chaque recommandation relative à l'évaluation. Le chef de l'évaluation (personne-contact) devrait préparer un plan d'action à soumettre à l'approbation du représentant dans le pays. Le plan d'action approuvé devrait devenir un appendice du rapport d'évaluation avant qu'il ne soit transmis pour archivage au siège de New York. Lorsqu'une recommandation sur l'évaluation suggère qu'un autre organisme entre en scène, l'UNICEF devrait solliciter une réponse de cet organisme. Pour faciliter les plans d'action à la suite des rapports d'évaluation, les constatations, enseignements et recommandations établis à titre provisoire devraient être soumis à une critique dans un atelier final auquel participent tous les protagonistes principaux.

En plus de ces recommandations principales, nous faisons les suggestions suivantes :

Formule et contenu généraux

1. Les rapports d'évaluation devraient être limités à une longueur plus ou moins standard, disons 50 pages + le résumé analytique + les appendices.
2. Les rapports d'évaluation devraient discuter de la qualité des données et expliquer les effets des contraintes imposées par l'impossibilité de joindre les gens, les départs à la retraite et les démissions, et l'absence de réponses.
3. Un résumé analytique du rapport d'évaluation devrait être traduit dans la/les langue(s) locale(s). Cela devrait faire partie du cahier des charges et inscrit comme il convient au budget.

Procédures générales

Nous suggérons que le Bureau des évaluations du siège de New York et les bureaux régionaux mettent au point un plan d'évaluation progressif de trois ans en conjonction avec les bureaux de pays pour coordonner l'activité d'évaluation et garantir une couverture suffisante aux questions les plus importantes. Ce plan devrait être remis à jour chaque année.

L'UNICEF a besoin de trouver une façon de tenir les bureaux de pays constamment au courant des évaluations entamées et achevées dans d'autres bureaux.

Nous suggérons que le texte complet des évaluations menées à terme pendant l'année en cours et les trois années précédentes soit accessible sur Internet dans le format Adobe, et que

les mots clés en facilitent la recherche et l'accès.²⁰ Nous croyons que, pour faciliter l'ouverture et inciter à la qualité, toutes les évaluations de l'UNICEF ayant coûté plus de, disons, 10 000 dollars, soient disponibles au public sur le site Internet ouvert.

1. L'UNICEF devrait exiger qu'une copie électronique de tous les cahiers des charges soient soumises au Bureau régional et au siège avant qu'une évaluation ne soit commandée, ainsi qu'une copie électronique du rapport d'évaluation après coup.
2. Pour s'assurer que les normes de qualité minimum sont respectées, l'UNICEF devrait envisager d'établir des conditions plus strictes pour la gestion, l'évaluation et la clôture des missions d'experts sectoriels, surtout lors de la phase de conception de la recherche sur l'évaluation. Ceux-ci devraient être reconnus dans le rapport final.

²⁰ Agence canadienne pour le développement international, op. cit., p. 48. « Les résumés analytiques sur l'évaluation, aussi utiles qu'ils soient pour alerter une large section du personnel sur l'existence d'un rapport d'évaluation, ne peuvent pas combler le fossé de la compréhension. Les évaluations sont complexes et demandent une lecture en profondeur si l'on veut que le projet ou le programme soient compris et les enseignements bien assimilés. »

RESUMEN EJECUTIVO

La calidad de las evaluaciones respaldadas por las oficinas del UNICEF en los países, 2000-2001

Propósito

Este estudio evalúa la calidad de las evaluaciones del UNICEF llevadas a cabo y comisionadas por sus oficinas en los países durante el período de 2000-2001. Ésta es la segunda revisión de la calidad de la evaluación que hace el UNICEF en esta década. Le siguen también varias revisiones temáticas, incluida una revisión de las evaluaciones educativas respaldadas por el UNICEF²¹.

Los objetivos de este estudio son valorar la calidad de las evaluaciones respaldadas por las oficinas del UNICEF en los países, ver si ha habido algún progreso desde la última revisión y recomendar el modo de mejorar la calidad.

Contexto

El UNICEF tiene una estructura descentralizada de rendición de cuentas para la evaluación²². Las oficinas del UNICEF en los países llevan a cabo evaluaciones de programas y proyectos por razones de supervisión y aprendizaje. Los informes resultantes son la materia de este estudio. Aunque la Oficina de Regiones y Evaluación del UNICEF en la sede de Nueva York desempeña papeles importantes, su trabajo queda fuera del alcance de este estudio.

Criterios de calidad

En este estudio se aplicaron 30 criterios de calidad de la evaluación. Estos criterios abarcan la calidad de la evaluación en todas las etapas del ciclo del proyecto de evaluación, desde el diseño hasta su aplicación. Aparecen listados en el Apéndice 1, junto con un resumen estadístico de los resultados de la revisión de la calidad en cada caso. Los criterios se identificaron a partir de las normas de asociaciones de evaluación profesional, publicadas, de fuentes textuales²³, especialmente de anteriores revisiones de calidad de la evaluación hechas por el UNICEF²⁴, la Agencia de los Estados Unidos para el Desarrollo Internacional (en inglés, USAID)²⁵, y la Agencia Canadiense para el Desarrollo Internacional (en inglés, CIDA)²⁶, y de discusiones entre los evaluadores profesionales en el UNICEF y en el equipo asesor.

²¹ Chapman, D. Revisión primaria de 185 evaluaciones.

²² UNICEF "Report on the Evaluation Function in the Context of the MTSP", E/ECEF/200210.

²³ Véase, por ejemplo, Thomas Black, *Evaluating Social Science Research: An Introduction*, Londres, Sage, 1993.

²⁴ C.G. Victora, et al. *A Systematic Review of UNICEF-Supported Evaluations and Studies, 1992-1993*, UNICEF, Nueva York, Documento de trabajo No. 3, junio de 1995.

²⁵ Centro para la Información y Evaluación del Desarrollo, Agencia de Estados Unidos para el Desarrollo Internacional, revisión de la calidad de las evaluaciones de la A.I.D.: años 1987 y 1988. Documento ocasional No. 19, USAID, Washington, D.C., mayo de 1989. También, Plan de acción: mejora de la evaluación central en la A.I.D., 1991.

²⁶ Agencia Canadiense para el Desarrollo Internacional, un estudio de la calidad de las evaluaciones bilaterales de la CIDA: 1988-89 y 1989-90, Ottawa, junio de 1991 (por el Dr. Kenneth Watson).

Se usó un cuestionario de 23 preguntas, algunas de ellas estructuradas y algunas de final abierto, para recopilar información de las oficinas del UNICEF en los países (Apéndice 2).

Alcance, limitaciones y metodología

Se revisó una muestra aleatoria del 50% de los informes de evaluación del UNICEF que han presentado las oficinas en los países a la sede de Nueva York en los años 2000 y 2001. Se halló que, de esos 97 informes, 75 eran evaluaciones, estrictamente definidas. En aproximadamente un tercio de los informes, revisamos también la calidad de los términos de referencia.

Además, se llevó a cabo una encuesta del personal en una muestra de un 50% de las oficinas del UNICEF en los países [47]. La respuesta se acercó al 100%.

Finalmente, el equipo de valuación examinó también la calidad de 31 términos de referencia para nuestra muestra de evaluaciones –aproximadamente un tercio– todas las cuales estaban disponibles en la sede de Nueva York para esta serie de estudios.

Los asesores de la calidad eran independientes. La revisión se comisionó mediante una oferta competitiva, y ningún revisor tuvo participación en ninguna de las evaluaciones. El diseño de investigación para este estudio no fue sometido a revisión ética. Sin embargo, ya que el trabajo se hizo enteramente a partir de informes del UNICEF ya existentes, y de información ofrecida por el personal del UNICEF en el curso de sus deberes profesionales, no se esperaba que surgiera ningún problema de consentimiento informado y competente. El único problema ético fue la protección de la confidencialidad del informante durante este trabajo.

Este estudio tuvo dos limitaciones significativas. Primero, estaba basado solamente en documentos, entrevistas en la sede de Nueva York, y en un cuestionario escrito para las oficinas en los países. No se realizaron visitas a estas oficinas, y no se interrogó en profundidad al personal de las oficinas en los países. Los procedimientos de evaluación no se observaron directamente, ni se revisaron los expedientes. No se entrevistaron a asociados ni usuarios de las evaluaciones del UNICEF. Nuestros juicios se basaron solamente en la calidad de los informes de evaluación y en sus términos de referencia escritos. En consecuencia, dispusimos de una escueta base para nuestras recomendaciones acerca de cómo podría mejorarse la calidad de la evaluación, y por tanto cualquier tipo de recomendaciones sobre cómo una mejor gestión de evaluación podría mejorar la calidad debe tener en cuenta este matiz.

La segunda limitación importante de este estudio fue que los informes de evaluación examinados se escogieron entre aquellos que las oficinas en los países sometieron a la sede de Nueva York, y no se presentaron todas las evaluaciones respaldadas por las oficinas en los países. Esperamos que los informes deficientes tengan menos probabilidades de ser sometidos a la sede. Los que respondieron al cuestionario indicaron que sus oficinas habían llevado a cabo aproximadamente 75 evaluaciones durante 2002 o 2001 que no se habían presentado a la sede de Nueva York. Sin embargo, puede haber habido más. Por otra parte, algunos informes no presentados puede que no hayan sido evaluaciones conforme a nuestra definición. El lector debe tener en mente que la presentación selectiva de informes de evaluación por las oficinas en los países puede haber dado lugar a opiniones prejuiciadas en nuestra muestra.

Constatación

Descripción general

Se encontró que las evaluaciones del UNICEF no son de calidad uniforme. Aproximadamente una de cada cinco es excelente, pero el tercio peor es lo bastante deficiente para constituir un problema.

Aspectos mejores y peores

Los cinco aspectos cualitativos en los cuales quedaron mejor las evaluaciones del UNICEF fueron:

- Los objetivos de muchas evaluaciones, y las preguntas a responder, con frecuencia fueron plena y claramente formuladas.
- Muchos informes de evaluación fueron claros, transparentes y fácilmente accesibles al lector. Los mejores fueron concisos, bien organizados y lógicos, con textos claramente escritos, respaldados por tablas, cifras y encabezamientos descriptivos, y presididos por un resumen ejecutivo.
- Los objetivos de la evaluación, y de las preguntas a responder, a menudo reflejaron la misión y el enfoque del UNICEF a la programación, incluida la protección de los derechos de los niños y niñas, la promoción de su bienestar y la igualdad de género (siendo el género el más débil de estos aspectos).
- Las recomendaciones con frecuencia estaban bien basadas en evidencias y análisis.
- La información cualitativa y cuantitativa reunida por muchas evaluaciones fue, en conjunto, adecuada para responder las preguntas de la evaluación.

Los cinco criterios de calidad en los cuales salieron peor las evaluaciones del UNICEF fueron:

- Los costos no estaban bien descritos, y rara vez se compararon con los resultados.
- Con frecuencia los “productos” del programa o proyecto no se describieron o se midieron correctamente y, con este eslabón perdido, se interrumpió la cadena causal de actividades a efectos directos.
- Rara vez se llevó a cabo una revisión ética en la etapa del diseño de investigación, y el tópico de la ética de la investigación rara vez se abordó en los informes. Es vital, por supuesto, que el diseño de la evaluación sea ético y que incluya salvaguardas éticas donde esto se requiera, entre ellas la protección de la dignidad, los derechos y el bienestar de los seres humanos, particularmente de los menores de edad, y el respeto por los valores de la comunidad beneficiaria. No tenemos ninguna opinión respecto a si hubo algunos problemas éticos en la investigación, tales como capacidad o consentimiento informado, pero simplemente notamos que las evaluaciones rara vez abordaron el tema. Los evaluadores rara vez hicieron una declaración acerca de cómo su objetividad e independencia quedaban garantizadas.
- Las evaluaciones fueron por lo general cortas de miras. Con frecuencia no se describió el grado hasta el cual el proyecto, el programa o la iniciativa podría reproducirse en otros contextos.

- Con frecuencia las lecciones aprendidas no se generalizaron más allá de la intervención inmediata que se evaluaba para indicar la mayor pertinencia que podría haber para el UNICEF.

Prioridades

Para mejorar la calidad de sus evaluaciones, el UNICEF debe concentrarse en el riesgo o la excelencia, o, por supuesto, en ambas cosas, si pueden movilizarse suficientes recursos.

1era. opción: Maximizar la influencia del UNICEF concentrando los empeños evaluativos en producir un número relativamente pequeño de excelentes evaluaciones de estrategias de intervención en esferas vitales para estas intervenciones, y con amplias posibilidades de reproducción.

2da. opción: Minimizar el riesgo aumentando la calidad del tercio más deficiente de las evaluaciones del UNICEF hasta alcanzar un mínimo de normas profesionales.

Las herramientas para mejorar el tercio peor de las evaluaciones deberían ser adecuadamente sencillas. Entre ellas podrían incluirse marcos genéricos para los términos de referencia de la evaluación (incluidas la normas de los procesos, el calendario y los presupuestos), y los índices regulares para los dos tipos de evaluaciones [evaluaciones de desempeño y evaluaciones de estrategias de intervención].

En el otro extremo del espectro de calidad, el problema y la respuesta adecuada son bastante diferentes. El UNICEF ha producido muchas evaluaciones que son buenas pero no excelentes. Alcanzar la excelencia, a partir de esta base de un buen trabajo, sería más difícil y costoso que la opción de la “norma mínima” descrita anteriormente. Tres cosas serían importantes:

- Emplear a los mejores profesionales de la evaluación, que probablemente resultarían muchísimo más caros de lo que algunas oficinas en los países creen que pueden costear.
- Insistir en rigurosos diseños de investigación que sean mucho menos impresionistas de lo que con frecuencia han sido las evaluaciones del UNICEF.
- Crear metodologías para evaluaciones de varias estrategias para intervenciones basadas en derechos. Las dimensiones culturales, políticas y jurídicas de las intervenciones basadas en derechos hacen su evaluación particularmente difícil, y el UNICEF con frecuencia estaría abriendo un nuevo camino en la metodología de la evaluación.

Alcanzar la excelencia exigiría gerentes muy adiestrados en evaluación, así como asesores altamente calificados y mayores presupuestos para obtener una recolección y análisis de datos primarios más completos. Todo esto suscita dudas de cuánto el UNICEF puede y quiere pagar por las evaluaciones. Estos problemas quedan fuera del alcance de este estudio. Quizás el único modo de alcanzar la excelencia dentro de un presupuesto costeable consiste en hacer cada año un número limitado de evaluaciones claramente pertinentes de estrategias de intervención en algunos asuntos claves, en las cuales participen varias oficinas en los países. El diseño y la coordinación de tales evaluaciones podrían exigir un liderazgo de las regiones y de la sede del UNICEF.

Otra aproximación a la excelencia, dentro de las limitaciones de la economía, es llevar a cabo más evaluaciones de estrategias de intervención conjuntamente con otras agencias para el desarrollo, incluidas las asociadas en el Marco Conjunto de Cooperación para el Desarrollo del Sistema de Naciones Unidas (en inglés, UNDAF), los bancos para el desarrollo multilateral y las principales agencias bilaterales. Los casos en que esto se hizo en evaluaciones de los resultados de intervenciones financiadas conjuntamente tendieron a ser de mejor calidad que los empeños hechos solamente por el UNICEF. Puesto que muchas esferas de derechos son de interés común, podría ser posible instituir una serie de evaluaciones conjuntas de estrategias de intervención en esferas claves con uno o más asociados entre los organismos internacionales.

Aumento de la capacidad del asociado nacional

Las sugerencias que apuntamos más arriba se relacionan con la propia capacidad y desempeño de evaluación del UNICEF. No mejorarán la capacidad de evaluación de las agencias del país anfitrión. Los intentos de mejorar la capacidad local mediante la participación de unos cuantos miembros del personal o de asesores locales, o de ambos, en las evaluaciones del UNICEF en el desempeño de papeles menores puede no resultar muy eficaz. La mayoría de los países en desarrollo necesitan más asistencia estructural para mejorar sus capacidades de evaluación, tales como la formación de una asociación nacional de evaluación afiliada con una asociación profesional internacional. Luego, otras agencias internacionales están interesadas y participan en aumentar las capacidades de evaluación de asociados nacionales, y el UNICEF debe llevar a cabo proyectos conjuntos para aumentar la capacidad dondequiera que sea posible.

Recomendaciones principales

Nuestra recomendación general es, por supuesto, que se haga todo lo posible para mejorar la calidad de las evaluaciones del UNICEF. Sin embargo, somos conscientes de que este estudio es sólo parcialmente adecuado como una base para las recomendaciones sobre cómo llevarlo a cabo. El mismo no abarca todas las evaluaciones del UNICEF, ni examina sus recursos, sistemas y prácticas de evaluación. Dicho eso, hacemos las siguientes recomendaciones principales:

1. La Oficina de Evaluación del UNICEF debería formular un plan de acción para el mejoramiento de la calidad de la evaluación en respuesta a este estudio.

Este estudio debería ser complementado por un estudio de los recursos y la organización de la evaluación del UNICEF. Dada la persistencia de los mismos problemas de calidad durante un largo período de tiempo, podrían necesitarse algunos cambios sistémicos. Pendiente del resultado de tal estudio complementario, debería considerarse la necesidad de equilibrar el sistema de evaluación descentralizado del UNICEF y basado en las oficinas en los países, con requisitos más severos para revisar los diseños de investigación de la evaluación fuera de la oficina en el país donde se inicia y en el momento en que se formulan los términos de referencia. Una mejora en los diseños de investigación es probablemente la única cuestión que tiene más posibilidades de mejorar la calidad de las evaluaciones del UNICEF. Cada estudio de evaluación debería tener una revisión y una comprobación metodológica hecha por un experto en la materia fuera de la oficina en el

país. Para cada evaluación presupuestada en más de 25.000 dólares, esta revisión debería estar basada en el marco de una evaluación completa²⁷.

2. Puesto que virtualmente todas las evaluaciones del UNICEF involucraran a personas, las evaluaciones de la oficina en el país deberían estar sujetas a requisitos más severos de revisión ética antes de llevarse a cabo. Sugerimos que la Oficina de Evaluación del UNICEF enuncie una política de revisión de la ética investigativa de la evaluación, y que se establezca un sistema adecuado a la misma. Esto incluiría una política sobre la revisión ética de los términos de referencia, la competencia del sujeto y el consentimiento informado, y una política sobre la información de acontecimientos adversos. No creemos que es suficiente para la Oficina de Evaluación depender de otros requisitos más estrictos para la revisión ética antes de su ejecución. Bastarían los enunciados políticos del UNICEF, a este respecto, o el de la Nota Técnica²⁸ existente. Además, cada informe de evaluación debería contener una declaración de cómo se garantizó la objetividad e independencia de los evaluadores.
3. Puesto que los informes de evaluación de la oficina en el país siguen mostrando muchas de las deficiencias que se encontraron en 1995 respecto a los informes de evaluación de principios de la década del noventa, y pese a un adelanto mostrado en la política, las directrices y las notas técnicas de la evaluación, recomendamos que ciertas cosas sean un *contenido obligatorio* en todos los estudios de evaluación (reforzado este requisito por índices modelos para los informes de evaluación). Los informes de evaluación deben seguir un formato regular, a menos que exista una buena razón para variarlo²⁹. Los términos de referencia del UNICEF para los estudios de evaluación deberían incluir el bosquejo de un índice de materias para el informe de evaluación, semejantes a los modelos que aparecen en los apéndices 3 y 4. El contenido regular debería incluir lo siguiente:
 - Debería haber una descripción clara de lo que se evalúa, es decir, suficiente información de antecedentes y contexto para permitirle a un lector no familiarizado con el país y el programa entenderlo plenamente, y describir explícitamente la pertinencia y replicabilidad en otros contextos del proyecto o programa y su evolución. Todas las evaluaciones del UNICEF de estrategia de intervención deberían consignar la más amplia pertinencia de las lecciones aprendidas, y la replicabilidad y ampliación a escala de los aspectos exitosos del proyecto o del programa. Debería haber una reseña completa de la intervención y su contexto, un análisis de atribución donde sea necesario, y una reflexión sobre la replicabilidad, ampliación de la escala, sostenibilidad y aspectos ambientales. Debería enunciarse claramente si el estudio es el análisis de una situación, una evaluación de estrategias de intervención alternativas o una evaluación del UNICEF (o ambas cosas), o el desempeño anterior de su asociado en un programa o proyecto.

²⁷ Un “marco de evaluación” comprende un conciso perfil del proyecto o programa a evaluar, incluido un modelo lógico que sostiene su razón de ser (un análisis del marco lógico u otro modelo apropiado), una lista de las preguntas que deberá responder el equipo de evaluación y una descripción de la metodología adecuada al tipo de estudio [evaluación de la estrategia o evaluación de los resultados] y a las preguntas.

²⁸ Oficina de Evaluación del UNICEF, Nota Técnica No. 1, “Children Participating in Research, Monitoring and Evaluation – Ethics and Your Responsibilities as a Manager”, abril de 2002.

²⁹ La normación del contenido/formato del principal informe de evaluación no excluiría productos en otros formatos para varias obras de divulgación.

- Un análisis de costos y, donde corresponda, de eficiencia, debería ser parte de todas las evaluaciones. El UNICEF debería tener normas sobre cómo calcular los costos totales de un proyecto o programa, incluido el tiempo del personal del UNICEF, los contratos, los costos de asociados y los costos de participantes.
- Debería haber una medición detallada de los productos del programa o proyecto. Para posibilitar esto, los contratos o acuerdos del UNICEF con los agentes encargados de aplicarlos deberían basarse en el desempeño y el producto. Esta base de acuerdos diáfanos de desempeño-producto es esencial en un enfoque basado en los resultados para una buena gestión, y para posibilitar una buena evaluación del desempeño.
- Si se espera que el proyecto o el programa resulte sostenido, con financiación del UNICEF o sin ella, entonces los mecanismos para garantizar esto deberían explicarse. Si las obligaciones de terceros son necesarias, deberían describirse. La autosostenibilidad mediante la recuperación de costos siempre debería ser una alternativa de sostenibilidad valorada.

La oficina del UNICEF en el país debería registrar su dictamen sobre cada evaluación que recomiende. El gerente de evaluación (coordinador) debería preparar un plan de acción sujeto a la aprobación del representante en el país. El plan de acción aprobado debería convertirse en un apéndice del informe de evaluación antes de someterlo a la sede en Nueva York para ser archivado. En el caso de que una recomendación sobre la evaluación requiera la intervención de otra agencia, el UNICEF debería pedirle a esa agencia una respuesta. Para facilitar los planes de acción que siguen a los informes de evaluación, los hallazgos provisionales, las lecciones y las recomendaciones deben estar sujetos a confrontación en un “curso práctico de salida” que involucre a todas las principales partes interesadas.

Además de estas principales recomendaciones, hacemos las siguientes sugerencias:

Formato general y contenido

1. Los informes de evaluación deberían limitarse más o menos a una extensión regular de presentación, digamos 50 páginas, más el resumen ejecutivo y los apéndices.
2. Los informes de evaluación deberían entrar a debatir la calidad de la información, y explicar los efectos de limitaciones de alcance, desgaste o falta de respuesta, o una combinación de estos factores.
3. Un resumen ejecutivo del informe de evaluación debería traducirse al idioma o idiomas locales. Este debería ser parte de los términos de referencia y presupuestarse adecuadamente.

Procedimientos generales

Sugerimos que la Oficina de Evaluación de la sede en Nueva York y las oficinas regionales creen un plan de evaluación de tres años en conjunción con las oficinas en los países a fin de

coordinar la actividad evaluativa y garantizar la adecuada cobertura de los problemas más importantes. El plan debería actualizarse todos los años.

El UNICEF necesita algunos medios de actualizar frecuentemente la información de la oficina en el país respecto a qué evaluaciones han comenzado y concluido en otras oficinas nacionales.

Sugerimos que el texto completo de las evaluaciones concluidas del año en curso y de los dos años anteriores esté disponible en Internet en formato de Adobe, y con palabras claves para facilitar la búsqueda y el acceso³⁰. Creemos que, para facilitar la apertura y brindar un incentivo que mejore la calidad, todas las evaluaciones del UNICEF que cuesten más de 10.000 dólares, por ejemplo, deberían ser accesibles al público en el portal abierto de la Internet.

1. El UNICEF debería exigir que una copia electrónica de todos los términos de referencia deba presentarse a la Oficina Regional de la sede en Nueva York antes de que se contrate una evaluación y una copia electrónica del informe de la evaluación que sigue.
2. Para garantizar que se cumplen un mínimo de normas de calidad, el UNICEF debería considerar el instituir más requisitos para la gestión, la evaluación y la aprobación de los expertos sectoriales, especialmente en la etapa del diseño de la investigación evaluativa. La aprobación de rigor debería reconocerse en el informe final.

³⁰ *Canadian International Development Agency*; op. cit., p. 48. “Los resúmenes de evaluación, tan útiles como son en alertar a una amplia muestra representativa de personal sobre la existencia de un informe de evaluación, no pueden salvar las deficiencias de comprensión. Las evaluaciones son complejas y exigen una lectura en profundidad si el proyecto o programa ha de entenderse y las lecciones captarse”.

1. INTRODUCTION

1.1. Objectives

This is a study of the quality of UNICEF evaluations conducted and commissioned by its country offices in 2000-2001. It is the second such review of evaluation quality. The first, done in 1995, examined evaluations completed in 1992-93.⁴

The objectives of the study are to identify the level of quality, to see whether progress has been made since the last review, and to recommend how quality might be improved. (See Appendix 5, Terms of Reference.)

Context

UNICEF has a decentralised accountability structure for evaluation.² The UNICEF country offices undertake evaluations of programmes and projects for oversight and learning. The resulting reports are the subject of this study. Although the UNICEF Regions and Evaluation Office at New York headquarters also conduct evaluations, these were outside the scope of this study.

Country offices conduct most evaluations. In each country office, an evaluation focal point is accountable to the country representative, who reports annually to the regional director on evaluation findings. UNICEF country representatives ensure that adequate staff resources are dedicated to evaluation, that communication with government officials and other partners facilitates the evaluation process, and that evaluation findings inform the decision-making process. UNICEF representatives are responsible for the Integrated Monitoring and Evaluation Plan (IMEP) and the quality of the evaluations they support.

Each regional office has a monitoring and evaluation officer who coordinates evaluation work performed by the country offices and their own regional office.

Regional offices provide oversight and support for evaluations undertaken by the country offices. Regional offices also conduct thematic evaluations related to their regional strategies. The UNICEF regions have responsibilities to strengthen the monitoring and evaluation capacities of UNICEF country offices and their government counterparts through coordination, preparation of regional evaluation plans, technical assistance, and oversight of country office evaluations. Headquarters divisions also undertake evaluations relating to their areas of expertise. The evaluations by regions and headquarters divisions were not part of this study.

The Evaluation Office at headquarters provides functional leadership and overall management of the evaluation system, and conducts evaluations. The office reinforces UNICEF's evaluation capacity, with an emphasis on the requirements of country offices and capacity-building in countries; and maintains a database of evaluations and research studies. The office collaborates with other United Nations agencies to harmonise evaluation activities and guidelines. It also monitors and reviews the quality of UNICEF-sponsored evaluations, and, in that context, commissioned this study.

1.2. Methodology

Criteria of Quality

The criteria of evaluation quality applied in this study (see Box 1 and Appendix 1), and the questionnaire by which information was gathered from UNICEF country offices (Appendix 2), were identified from the published evaluation standards of professional evaluation associations; from literature³, especially previous reviews of evaluation quality by UNICEF⁴, USAID⁵ and CIDA⁶, and from discussion among the professional evaluators in UNICEF and within the consulting team.

Sample of Reports

The criteria of quality were applied to a 50% random sample of UNICEF evaluation reports submitted to New York Headquarters by the country offices in the years 2000 and 2001 (approximately 97 reports, of which 75 proved to be evaluations, strictly defined). The sample was drawn from a list of evaluations that had first been reviewed to sort the true evaluations from other types of research.³¹ As well, the assessment team examined the quality of 31 Terms of Reference for our sample of evaluations — about one third — all that were available at New York headquarters for this set of studies.

Assessment Procedures

Five consultants — all expert evaluators — assessed the quality of the UNICEF evaluations. (See Appendix 6.) . The team leader did the first five assessments to benchmark the standards of quality. Each evaluator reviewed a benchmark assessment, and completed an initial assessment of one evaluation that was discussed with the team leader in detail. The evaluators then proceeded to complete the set of assessments, discussing issues as they arose.

Reports were randomly assigned to assessors, except in so far as the language of the report was a constraint. Only one evaluator was competent to assess evaluation reports in Spanish, and one in French.

Assessors were engaged primarily for their expertise as evaluators and evaluation methodologists, but, in fact, the team did have wide experience in the main areas of UNICEF's work as well.

Survey of Country Offices

A questionnaire was sent to staff in the country offices — either the UNICEF country representative or the evaluation focal point in the office. The questionnaire was developed in conjunction with UNICEF headquarters staff, and tested with three country office focal points before it was finalised. (See Appendix 2.)

A 50% random sample of country offices received the survey. The reason for a sample was to minimise the response burden. By careful follow-up, the study team achieved virtually a 100% response. Only one country office declined to respond.³²

³¹ Research was considered an evaluation only if it examined an actual or potential investment or other intervention by UNICEF. By this measure, some research reports were misclassified as evaluations and were put aside.

³² One office declined to respond on the grounds that it had done no evaluations during 2000 and 2001, no staff currently in the office had been there during that time, and no current staff had evaluation experience.

BOX 1: CRITERIA OF EVALUATION REPORT QUALITY

General format of the evaluation report:

Was there a cover page and title? Was the report dated, and the author(s) named? Was there a table of contents? An Executive Summary? Were Terms of Reference attached?

Criteria of Quality

1. The project, programme, or initiative to be evaluated should be clearly described, including the logic of its links between activities, outputs, and outcomes.
2. The degree to which the project, programme, or initiative might be replicable in other contexts should be described.
3. The incremental contributions of UNICEF and other stakeholders, in cash or in kind, should be stated.
4. The context should be analysed in sufficient detail to identify external factors contributing to success or failure, and this analysis should be used to support informed judgements about what results may reasonably be attributed to UNICEF's intervention, and what to other factors.
5. The objectives of the evaluation, and the questions to be answered, should be stated fully and clearly.
6. The objectives of the evaluation, and the questions to be answered, should reflect UNICEF's mission and approach to programming, including protection of children's rights, promotion of their welfare, and gender equality.
7. The evaluation should use standard UNICEF evaluation criteria: relevance, efficiency, effectiveness, impact and sustainability.
8. The evaluation report adequately addresses the issue of sustainability.
9. The evaluation report adequately addresses the issue of relevance.
10. For evaluations of humanitarian response, in particular at the sector or whole country level, the evaluation should address coverage, coordination, coherence and protection.
11. The methodology of the evaluation should be practical and appropriate to the questions posed, while providing a complete and fair assessment.
12. The evaluation design should be ethical and include ethical safeguards where appropriate, including protection of the dignity, rights, and welfare of human subjects, particularly children, and respect for the values of the beneficiary community.
13. The evaluators should make a statement about how objectivity and independence were ensured.
14. The constraints of the evaluation, and the perspective from which the intervention is evaluated, should be clear so the reader can assess the validity of the evaluators' judgements.
15. The evaluation report should describe the kinds of UNICEF and other key stakeholders' follow-up action expected in regard to lessons and recommendations.
16. The information gathered, qualitative and quantitative, should, in aggregate, be adequate to answer the evaluation questions.
17. Sampling is used when appropriate and is designed to produce representative data.
18. The data collection instruments should be able to provide the measurements needed.
19. Information should be gathered from eligible persons not reached in addition to those reached.
20. The report should distinguish clearly between implementation factors [inputs, process milestones, activities] and results [outputs and outcomes/impacts].
21. Inputs [costs/resources] should be measured.
22. Outputs [what is actually produced by the inputs] should be measured.
23. Outcomes/impacts [what the intervention seeks to achieve] should be measured or an appropriate rationale given why not.
24. Costs and results should be directly compared to as full an extent as possible.
25. Accountability for results [who is accountable for what, and their performance] should be appropriately analysed and reported.
26. Lessons learned should be generalised beyond the immediate intervention being evaluated to indicate what wider relevance to UNICEF there might be.
27. Recommendations should be firmly based on evidence and analysis.
28. Priorities for action should be made clear.
29. The evaluation report should be complete, with all of the necessary elements of a good report present.
30. The evaluation report should be clear, transparent and easily accessible to the reader. It should be concise, well organised and logical. The text should be clearly written, supported by tables, figures and descriptive headings, and led by an executive summary.

1.3. Scope and Limitations

This study had two significant limitations. First, it was based solely on documents, interviews at New York headquarters, and a written questionnaire for country offices. Country offices were not visited, and country office staff not interviewed in depth. Evaluation procedures were not observed directly, nor files reviewed. Partners and users of UNICEF evaluations were not interviewed.

These limited our judgement of the quality of evaluations to a degree, and moreover, limited the confidence with which we have made recommendations on how the evaluation system might be changed to produce better quality reports.

Another important limitation of this study was that the evaluation reports examined were sampled from those submitted to New York headquarters by the country offices, and not all evaluations are submitted. We expect that poor reports are less likely to be submitted to headquarters, but we do not know this for sure. Table 2.1.1 shows the numbers of evaluations in the set sampled. Approximately 98 were in the initial sample, of which 75 were deemed, on close examination, to be true evaluations.

Respondents to the questionnaire indicated that their offices had completed approximately 75 evaluations during 2000 or 2001 that had not been submitted to New York headquarters (roughly as many as we reviewed). However, the number of reports not submitted was not verified. Some of the 75 may not have been evaluations by our definition. Nevertheless, it is a substantial number, and the possibility of bias in our sample remains.

1.4. Research Ethics

The quality assessors were independent. The review was commissioned by competitive tender, and no reviewers have been involved with any of the evaluations or engaged in any other way with UNICEF.

UNICEF country offices were assured anonymity in responding to the questionnaire. They sent the response directly to the consultants. No identifiers are included in this report, and none are attached to data we supplied to UNICEF headquarters. This ensured confidentiality to the country office. However, we noticed that many of the country offices required the respondent to clear the responses to the questionnaire with country office management.

The research design for this study was not subject to ethical review. However, since the work was done entirely from existing UNICEF reports, and from information proffered by UNICEF staff in the course of their professional duties, no issues of informed and competent consent were expected to arise. The only ethical issue was protection of informant confidentiality (see above).

1.5. Previous Reviews of Evaluation Quality

In 1995, UNICEF conducted a review of the evaluations and studies that it had supported in 1992-1993.⁴ This followed similar studies of evaluation quality that had been conducted by USAID⁵, and the Canadian International Development Agency.³³

³³ K. Watson, "A Study of the Quality of CIDA's Bilateral Evaluations, 1988-89 and 1989-90", Ottawa, June 1991.

The UNICEF study of evaluation quality concluded:

“Three percent of the reports were so inadequate that the reviewer was unable to give a rating; 29% were judged to be poor; 28% fair; 25% good; and 15% excellent. When the various report sections were considered, the executive summary received the lowest scores [56% of reports did not have one at all], closely followed by the recommendations, which were felt to be too general or unrelated to the study findings. Most reports were also criticised for lack of information on methodology as well as inadequate background data on the population, project or programme. The presentation of findings was in general the most satisfactory section. The reviewers felt that the most common reasons for inadequate reports might have been a lack of communication between researchers and UNICEF officers, lack of foresight (no baseline data, insufficient time or resource allocation to research) or inadequate competence of the investigators in the field under study.”³⁴

Recommendations of the study included:

- Improve the data base of evaluation reports
- Monitor the costs of evaluations
- Develop mechanisms to review research designs to ensure good quality
- Make capacity-building components of commissioned research mandatory
- Establish mechanisms to ensure appropriate action on recommendations
- Promote greater involvement of UNICEF officers in the research
- Develop training packages to address the most common deficiencies of evaluations
- Periodically repeat the review of evaluation quality.

³⁴ Victora, op. cit. p. vi.

2. UNICEF EVALUATION ACTIVITY

2.1. Country Office Evaluation Activity

One hundred and seventy-two evaluation reports were completed in 2000 or 2001 and held by UNICEF New York headquarters at the time of this study. Approximately 40% of the country offices we surveyed told us that the office had undertaken one or more evaluations during this period that had not been submitted to NYHQ. The number of evaluations not submitted, extrapolated from our sample to all country offices, was approximately 150.

Table 2.1.1: UNICEF Evaluations from 2000 and 2001 Received by NYHQ before December 2002

	Evaluations	Studies	Surveys
Region			
CEE/CIS	25	37	37
EAPRO	31	36	30
ESARO	28	38	30
MENA	20	18	19
ROSA	10	4	6
TACRO	25	62	21
WCARO	33	42	13
Uncoded	2	8	2
Total	174	245	158
MTSP			
Child Protection	12	49	12
ECD	13	13	14
Girls' Education	7	9	1
HIV/AIDS	17	19	20
Immunisation	9	4	7
Other	116	151	104
Total	174	245	158
Language			
English	114	119	117
French	36	45	16
Spanish	17	55	13
Other	7	26	12
Total	174	245	158

Despite this high level of activity, three quarters of the country offices in our survey told us that there were areas and topics globally or in each country that UNICEF has not evaluated sufficiently.³⁵

Recommendation [2.1.1] NYHQ Evaluation Office and Regional Offices should develop a three-year rolling evaluation plan, in conjunction with the country offices, to coordinate evaluation activity and ensure adequate coverage of key issues. The plan should be updated annually.

³⁵ Appendix 2, Question 3.

Approximately 40% of country offices report that they completed evaluations during 2000 or 2001 that were not submitted to NYHQ. The number of evaluations not submitted was approximately 75 per year.³⁶

One quarter of the country offices believe that the overall level of evaluation activity supported by them in 2000 and 2001 was too low, and 75% think it was about right. No one thinks that too much evaluation activity is being supported.³⁷

2.2. Accessibility and Use of Evaluations

Three quarters of the respondents to our survey believe that evaluations undertaken by other country offices are “only accessible with some difficulty”. Nevertheless, most (85%) claim to have read at least one evaluation from another country office in the past year, and half claim to have read one in the past quarter. Almost one in five respondents claims to have read one in the past month.³⁸ This indicates substantial use of the UNICEF evaluations, not just by the commissioning country office but by other country offices as well. Nevertheless, comments by the respondents to our survey, and our own observations, indicate unresolved issues in the following areas:

Awareness

Recommendation [2.2.1] UNICEF needs some means of frequently updating country office awareness of what evaluations are being started and completed in other country offices.

Access

Recommendation [2.2.2] The full text of completed evaluations from the current year and two previous years should be accessible on the Internet in Adobe format, and key worded for easy search and access.¹⁰

Transparency

Recommendation [2.2.3] Every UNICEF evaluation should be written for a broad audience within the organisation, giving sufficient background and context to a reader unfamiliar with the country and programme, and explicitly describing the international relevance of points made in the evaluation.

Openness

Recommendation [2.2.4] To facilitate openness and provide incentive to improve quality, all UNICEF evaluations that cost more than \$10,000 should be available to the public on the open website.

³⁶ Appendix 2, Question 1.

³⁷ Ibid, Question 2.

³⁸ Ibid, Question 5.

3. QUALITY OF EVALUATION DESIGN

3.1. Types of Evaluations

Three types of research were represented among the UNICEF reports we reviewed. The first type was a situation analysis. These reports were not evaluations because they did not examine a UNICEF intervention, either prospectively or retrospectively. They may be important studies but they were not within our scope.

The second type was an evaluation of UNICEF or partner performance (Did the project or programme achieve its objectives?), and the third type is an evaluation of intervention strategy (Was the project or programme design successful and is it replicable?). We do not suggest that the UNICEF reports made this distinction, or that there were many pure examples of these types, however, we do suggest that it is a useful distinction. Some quality criteria apply much more to one than to the other type of evaluation. (See Table 3.1.1.) For example, it might be beside the point to ask whether an evaluation of the intervention strategy assesses whether management accountabilities are fulfilled — there might not be any specific accountabilities at the stage when intervention strategy is being evaluated. Conversely, a performance evaluation might legitimately be situation-specific. It might make little sense to ask about replicability or scalability or alternative approaches to a problem when the focus is on the success (or otherwise) of a specific implementation experience.

The previous review of the quality of UNICEF evaluations observed that about 20% of reports were misclassified in the Evaluation Database.³⁹ There is something more important here than simple misclassification. The distinction between situation analyses, intervention strategy evaluation and service performance evaluation should be clearer in the design of UNICEF research. There is advantage in being clear what kind of evaluation is being designed and conducted. Of course, one evaluation might have both performance and strategy components. If so, we suggest that they be kept distinct within the work.

Performance evaluations and evaluations of intervention strategy should generally be conducted at different stages of a programme lifecycle, with different objectives and different designs. Of the two, we believe that evaluations of intervention strategy are likely to be more important to UNICEF because UNICEF's influence on rights is not mainly through efficient services delivery, but rather through knowledge and example.

The key to good evaluations of intervention strategy is a comparison of alternatives. The comparisons do not necessarily have to be rigorously experimental, with treatment and control groups, to be useful. It would be good if this were possible, but rights-based programming may not lend itself to this evaluation approach. Nevertheless, more modest quasi-experimental models are often available.

³⁹ Victora, op. cit., p.9.

Table 3.1.1: Key Aspects of Quality, by Type of Research

Quality Requirement	All Research	Situation Analysis	Strategy Evaluation	Performance Evaluation
STATE PROBLEM TO BE REMEDIED OR GOAL	√			
STATE ANSWERABLE QUESTIONS	√			
CONDUCT ETHICAL REVIEW OF THE RESEARCH DESIGN	√			
Convene independent team of competent researchers	√			
Allocate sufficient resources to complete the work well	√			
Use rigorous sample, structured and verified data collection	√			
Conduct thorough analysis, with challenge	√			
Provide complete, concise report of research and findings	√			
Review country needs and readiness [Baseline]				
Identify situation dynamics: Risks/Opportunities ⁴⁰		√		
Review capabilities and constraints		√		
Identify feasible scope/objectives for UNICEF action		√		
Agree with partners what needs to be accomplished		√		
Bottom line = understanding of needs, capabilities, risks, trends, constraints, perceptions, and preferences		√		
Identify alternative interventions [direct service, volunteer, cooperative, or market-based] on the basis of their relevance, logic, cost, risk and potential benefit				
Conduct theory-based experiments [pilots] to test link between proposed interventions and desired outcomes			√	
Focus on replicability of the action or “treatment” in more than one context (potential for scaling up)			√	
Identify priorities [bottlenecks, critical path, optimal sequencing, vital/less vital factors, etc.]			√	
Analyse potential sustainability			√	
Focus on cost-effectiveness [costs and outcomes]			√	
Feedback in policy-time			√	
Bottom line = (re)design and resource allocations			√	
Evaluate implementation and accountabilities, including output-based implementation agreements				
Always measure actual outputs				√
Sometimes (re)measure actual outcomes				√
Analyse costs and efficiency				√
Analyse reach, compared with targets				√
Observe whether performance is actually sustained				√
Feedback in real-time				√
Bottom line = accountability, lessons, action plans				√

⁴⁰ Objectives should be SMART – specific, measurable, achievable, relevant, and time-bound.

3.2. Quality of Terms of Reference

UNICEF NYHQ Evaluation Office had copies of thirty-one terms of reference for our sample of evaluations – about one third. There were approximately equal numbers of TORs for each of the years 2000 and 20001. Two thirds were available electronically, and one third in paper.

Recommendation [3.2.1] UNICEF should require that an electronic copy of the Terms of Reference should be submitted to the Region Office and NYHQ before an evaluation is contracted. This would require some adjustment to existing roles and accountabilities.

Country offices were evenly split between those that were generally happy with the quality of Terms of Reference and those that were not.

Table 3.2.1: Were evaluations based on high-quality Terms of Reference?

Response	Percent
Almost always	41%
Generally, but important exceptions	46%
Often not	13%

The format and content of the terms of reference varied greatly. Most gave a brief background description, one half to two pages, and stated the objectives of the evaluation. They generally described methodology, a work plan, and some description of the responsibilities of the consultants and the research products to be delivered. Sometimes the Terms of Reference contained a task/personnel analysis, or a task/budget analysis, but most often not. Products, milestone dates, and payments were sometimes linked, but generally not.

Terms of Reference varied in ways that did not seem related to whether they were situation analyses, performance evaluations, or evaluations of intervention strategy.

Recommendation [3.2.2] UNICEF Terms of Reference should include a model Table of Contents for the evaluation report, such as those shown in Appendices 3 and 4.

Recommendation [3.2.3] UNICEF should develop a standard electronic form with guidelines for Terms of Reference for evaluations.

The Terms of Reference, in addition to being unsystematic and having gaps, were generally skimpy. This was not clearly a function of the size of the evaluation. Terms of Reference for evaluations that involved several person months of consultant time were often not noticeably more thorough than those for evaluations that involved only a couple of weeks.

There was seldom reference to a logical framework or some other type of evaluation planning exercise. This seems to have contributed to methodological weaknesses, including insufficient attention to describing the project or programme to be evaluated.⁴¹

⁴¹ Victora, op. cit., p. 26 states, “Researchers frequently appeared to have limited knowledge of the specific situation or project/programme being assessed.”

Table 3.2.2: Were the initiative and its logic clearly described?⁴²

Score	Percent of Evaluations
Zero or 1	33%
2	25%
3 or 4	47%
All	100%

The evaluation research-planning phase (including preparation of Terms of Reference) is when the quality of the whole evaluation could be most affected for the better. Finding the resources and accepting the necessary disciplines to improve the review of evaluation designs is a major challenge for UNICEF.

Recommendation [3.2.4] Substantial evaluations, say those budgeted at more than US\$25,000, should begin with an explicit research design exercise to produce an “evaluation framework”, which should be referenced by the TORs for the subsequent evaluation.

3.3. Evaluation Methodologies

People who are not professional researchers may underestimate the importance of good methodology because they underestimate the possibility of drawing the wrong conclusions. This is particularly true of evaluations of intervention strategy, but to a lesser extent perhaps, also true of performance evaluations and situation analyses.

UNICEF evaluations varied in their methodologies but overall were impressionistic rather than rigorous. The most common methodology was unstructured discussion during short site visits between evaluators and persons delivering the programme. This approach relies a great deal on the personal skills of the interviewer⁴³ and it is difficult to achieve consistency when more than one interviewer is involved.⁴⁴ Contact with the treatment (target) group was frequently brief and superficial. More direct measurement of inputs, outputs, and outcomes through systematic data collection using well-structured instruments is needed rather than impressionistic reporting of hearsay. UNICEF is not alone in this weakness; both CIDA and USAID reviews of evaluation quality have reported similar findings.⁴⁵

The best UNICEF evaluation methodologies are excellent. For example, we ranked 17% of the evaluation reports as four on the 0-4 scale for having practical, complete and fair

⁴² Appendix 1, Criterion 1.

⁴³ Structured interview protocols can help, but are of limited use in many site-visit situations.

⁴⁴ The 1983-84 review of the quality of CIDA evaluations noted a similar problem: “The main problem lies in the collection and data analysis tools used. There should be more evaluators using interview grids or questionnaires, content analysis grids for the documents consulted and particularly the project files, and observation grids or other direct measurement tools to structure field visits. More than two thirds of evaluators do not mention using any tool, which does not help the accuracy of the analysis.” p. 17.

⁴⁵ CIDA, Area Coordination Group, “Review of Bilateral Evaluations 1981-83”, Vol. 5, pp 3-13: “The main problem (with CIDA evaluations) lies in the collection and analysis tools used. There should be more evaluators using grids or questionnaires, content analysis grids for the documents consulted and particularly for the project files, and observation grids or other direct measurement tools to structure field visits. More than two thirds of the evaluators do not mention using any tool, which does not help the accuracy of the analysis.”

methodologies.⁴⁶ However most methodologies were mediocre to poor, when described, and were frequently not described clearly. This mirrors the finding of the previous review of the quality of UNICEF evaluations that stated, “The instruments used were adequately documented in 36% of the reports, but the actual research procedures (including interviewer selection and training, data collection, data processing, quality control) were discussed in only one in five reports.”⁴⁴ The reviewers felt that the lack of methodological details often implied that insufficient attention was given to these issues.⁴⁷

Very few of the reports we examined contained a review of the literature, even a review of previous UNICEF studies. This gap is most serious in the case of evaluations of intervention strategy, but can be a significant problem for situation analyses and performance evaluations as well.

Sampling was seldom well done. The approach to sampling in half the reports we examined was rated zero or one on the 0-4 scale of quality.⁴⁸ Sampling frames, when they were described at all, were often ill specified. Where they were discussed, sampling proportions and response rates tended to be confused with each other. On the other hand, almost a third of UNICEF evaluations did well on sampling and about 10% were exemplary.⁴⁷

Many UNICEF evaluations did not use standard UNICEF evaluation criteria.⁴⁹ The quality score on this item was about half what it would have been had all evaluations used the standard criteria. (See Table 3.3.1 below.)

Table 3.3.1: Performance Scores on Evaluation Methodology Quality

Method	Actual Quality Score as a Percent of the Possible Score
Sampling ⁴⁸	39%
Use standard UNICEF criteria ⁴⁹	50%
Good data collection instruments ⁵⁰	52%
Constraints clearly described ⁵¹	54%
Programme logic clearly described ⁴²	57%
Methodology complete and fair ⁵¹	58%
Information adequate to questions ⁵²	61%

⁴⁶ Appendix 1, Criterion 11.

⁴⁷ Victora, op. cit., p.16. These findings are similar to those of the previous review of UNICEF evaluations, which stated: “Sampling schemes were inadequately described in two out of three reports that employed sampling. The most common problems included lack of representativeness and insufficient sample sizes. Lack of information on sampling often limited the interpretation of the report findings.”

⁴⁸ Appendix 1, Criterion 17.

⁴⁹ Ibid, Criterion 7.

⁵⁰ Appendix 1, Criterion 18.

⁵¹ Ibid, Criterion 14.

⁵² Ibid, Criterion 16.

In UNICEF evaluations of intervention strategy, perhaps the most serious methodological weakness was that explicit comparisons between alternative approaches were rare. In UNICEF performance evaluations, perhaps the most serious methodological weaknesses were the absence of a baseline⁵³ and a lack of measurable targets against which actual performance could be judged. Arguably, the latter are programme design flaws rather than evaluation flaws.

Recommendation [3.3.1] The Terms of Reference for each evaluation study should state whether the study is to contain a situation analysis, an evaluation of intervention strategy and/or a performance evaluation, and for each of these that it does contain, an appropriate methodology should be described.

Recommendation [3.3.2] Each evaluation study should have a methodological review and challenge by a peer in the country office, and, when the evaluation is important and difficult, by the Regional Office, or, when more appropriate, by someone in another CO. For each evaluation budgeted at over \$10,000, this review should be based on a full framework.⁷

Recommendation [3.3.3] The Regional Offices and/or NY HQ should provide sampling guidelines and consultation to evaluation managers in country offices.

⁵³ Victora, op. cit., p.12 had similar findings. "There were several instances in which baseline data were not collected, which precluded an adequate evaluation. A related problem was that often the evaluation exercise was planned with insufficient time or resources relative to the scale of the required task."

4. QUALITY OF EVALUATION MANAGEMENT

4.1. Purpose and Timing

Three quarters of the evaluation reports received an adequate or better score on clarity and completeness of objectives and issues. (See Table 4.1.1.)

Table 4.1.1: Clarity and Completeness of Objectives and Issues⁵⁴

Score	Percent
Zero or 1	23%
2	11%
3 or 4	61%

Most respondents to our survey believe that evaluations are generally timely. However, their comments indicate that many think of timeliness in terms of completing the evaluation contract on time, rather than doing the evaluation at the right time. For example, most performance evaluations are best done about one year into implementation so there can be mid-course corrections. Occasionally we saw this done well. However, most evaluations were scheduled when a funding decision was imminent. This is not necessarily optimal either from an "intervention strategy" or from a performance assessment point of view.

Table 4.1.2: Were evaluations timely in their findings and recommendations?

Response	Percent
Almost always	65%
Sometimes	26%
Seldom or never	7%

Recommendation [4.1.1] UNICEF should schedule evaluations of intervention strategy and performance evaluations at points where most can be learned, and/or when decisions depend on the evaluation (not only funding decisions).

4.2. Research Ethics

None of the evaluations we examined underwent "research ethics review" in the planning stage. The comments by some respondents to our survey of country offices indicated little knowledge of what research ethics entail. Not surprisingly then, the "ethics" quality criterion received the lowest performance score in our review,⁵⁵ mainly because the topic was seldom addressed in

⁵⁴ Appendix 1, Criterion 5.

⁵⁵ Ibid, Criterion 12.

the reports. Whether there were serious ethical issues is unknown. The previous general review⁴ of the quality of UNICEF evaluations did not address the issue.

Table 4.2.1: Quality of Ethical Safeguards⁵⁶ Reported by Evaluations

Score	Percent
Zero	67%
1	3%
2	5%
3 or 4	30%

About one quarter of respondents to our survey of country offices reported that ethical problems, such as protecting human subjects or respecting community values, had arisen during evaluations, but were almost always fully resolved. Another 7% said that ethical problems were occasionally unresolved, and 2% said that they were frequently unresolved.⁵⁷ We cannot verify these figures because UNICEF has no system for reporting adverse events that occur in association with its research.

Table 4.2.2: Did ethical problems, such as protecting human subjects or respecting community values, arise during the evaluation?

Response	Percent
No problems arose	65%
Problems arose but were almost always fully resolved	26%
Problems arose and were occasionally unresolved	7%
Problems arose and were frequently unresolved	2%

The UNICEF “Evaluation Technical Note”⁸ on the ethics of children’s participation in research and evaluations post-dates the reports that we examined, but there is no formal UNICEF policy on research ethics review, and no ethics section in the UNICEF *Evaluation Policy and Planning Manual*. This is a serious gap because informed consent to collect data in situations involving children in communities that may be culturally distinct from the researchers and disadvantaged in their own social context are important and complex. It is unusual today for academic research involving human subjects to proceed without formal ethical review by a Research Ethics Board. UNICEF’s standards should be equally high.

⁵⁶ Appendix 1, Criterion 12.

⁵⁷ Ibid, Criterion 10.

Objectivity

By and large, the UNICEF country offices think that the independence of evaluators is adequately protected. However, the tenor of the comments indicates that UNICEF staff may not be sufficiently aware of threats to objectivity.⁵⁸

Table 4.2.3: Was the independence of evaluators adequately protected?

Response	Percent
Almost always	83%
Generally, but with some important exceptions	15%
Frequently not	2%

We looked for assurance in the evaluation reports that objectivity had been adequately protected and often found little on the subject. Approximately 60% of reports received a zero or 1 on the 0-4 scale.

Table 4.2.4: Is there evidence in the evaluation reports that objectivity and independence were protected?⁵⁹

Score	Percent
Zero or 1	61%
2	15%
3 or 4	24%

One thing that gives us pause in regard to objectivity is that we very seldom found recommendations to terminate or not to re-fund a project/programme. Evaluators virtually always recommended incremental improvements and continued or expanded funding.

Recommendation [4.2.1] UNICEF should state an evaluation research ethics review policy, including a policy on subject competence and informed consent, and a policy on adverse event reporting.

⁵⁸ Appendix 2, Question 19.

⁵⁹ Appendix 1, Criterion 13.

Recommendation [4.2.2] UNICEF should institute a research ethics board, and the board should apply for accreditation, perhaps by the AAHPP.⁶⁰

Recommendation [4.2.3] Each evaluation report should contain a statement of how the objectivity and independence of the evaluators and their report were ensured.

4.3. Quality Control of Evaluations

Control of evaluations depends on several factors and mechanisms, including ethical review, adequate budgets, good Terms of Reference, skilled project managers, appropriate review and sign-off at key milestones, skilled research and good timing.

Adequate Budgets

Many of the evaluation reports that received good reviews were joint efforts by UNICEF and other development agencies. There may be many reasons for this, but one reason is probably that the total resources devoted to the evaluation were greater than UNICEF alone typically invests. The previous study of the quality of UNICEF evaluations found a substantial number of studies that did not provide value-for-money and “also noted the opposite situation: that is, studies that had achieved important results at a very low cost.”⁶¹ However, a CIDA review⁶² of the quality of its evaluations in 1991 concluded that much larger evaluation budgets than the typical \$US 7,000 UNICEF evaluation budget⁶³ “may not be adequate to produce more than a superficial review of the project”.⁶⁴ We note, however, that data on the costs of UNICEF evaluations are incomplete and do not take into account in-house resources or that of partners.

⁶⁰ The AAHRPP is a non-profit organization founded in May 2001 to provide voluntary accreditation services to research institutions regarding their human research protection programs. The founding organizations included the Association of American Medical Colleges, the Association of American Universities, and the Consortium of Social Science Associations. The AAHRPP piloted some trial accreditation examinations in November 2001, specifically, examining the 14 Institutional Review Boards of the National Institutes of Health. The goal of AAHRPPSM accreditation is to improve the systems that protect the rights and welfare of individuals who participate in research. In addition, accreditation can help to communicate to the public the strength of an organization's commitment to the protection of human research participants. It will also improve the overall quality of research by consistently applying high standards and practices, raising the global benchmark for human research protection. To help promote all of these goals, AAHRPP has adopted nine principles for accreditation of human research protection programmes. These nine principles serve as the foundation for the structure and content of AAHRPP's accreditation standards. The standards themselves are designed to help organizations consistently meet ethical principles and standards for protecting research participants, yet be flexible enough to account for the diverse institutional and cultural contexts in which research is conducted and reviewed. The structure of the accreditation standards owes much to the analysis offered by the Institute of Medicine in its report, "Preserving the Public Trust: Accreditation and Human Research Participant Protection Programs" (2001). AAHRPP's approach to voluntary accreditation incorporates five domains of a highly developed human research program: Organization, Research Review Unit, including IRBs, Investigator, Sponsor and Participant. The domains refer to different areas of responsibility that must be addressed in a human research protection programme. Meeting the requirements for all five domains is the responsibility of the entity seeking accreditation of its human research protection programme — the domains do not refer to separate persons or entities. Rather, the entity seeking accreditation must think of the domains as they relate to its human research protection programme.

⁶¹ Victora, op. cit., p.11.

⁶² CIDA, op. cit., 1991, Section 3.3.

⁶³ Victora, op. cit., Sub-section C.

⁶⁴ Ibid, p.36.

Recommendation [4.3.1] UNICEF should undertake a study of the costs of evaluations, and produce guidelines for appropriate budgeting of situation analyses, evaluations of intervention strategy, and performance evaluations. The option of fewer but better studies should be assessed. Evaluation budgets should be adequate to allow a thorough evaluation design, sufficient time in fieldwork to obtain good data, and adequate time for analysis and reflection.

The Skills of Evaluation Managers

Although we achieved virtually 100% response from country offices to our survey on the topic of evaluation quality, it took longer than we expected. One reason for this was that a significant number of staff who are evaluation focal points are new to the position, and do not have much experience managing evaluations.

If the country office evaluation managers have insufficient time, training, or experience for the job, or if there is significant conflict of interest in managing the evaluation of a project/programme that is one's own work, or the work of close colleagues, then some strengthening of evaluation support and/or direct evaluation activity regionally or at headquarters may be in order. We are inclined to think that most performance evaluations can be designed and managed locally and that most evaluations of intervention strategy should be cross-country. However, this requires study. The responsibility for evaluation in UNICEF may need to be more focused, and management of evaluations may need to be more professional, if their average quality is to improve. However, the general topic of how evaluation in UNICEF should be organised is beyond the scope of this study.

Recommendation [4.3.2] UNICEF should study its options in regard to the organisation of the evaluation function.

"Generic" Terms of Reference

A certain amount of standardisation would probably not affect the evaluations that are already on the path to excellence, but generic Terms of Reference may play a useful role for evaluations that would otherwise contain significant gaps. This does not mean that generic frameworks should be followed slavishly, or that they will enable an inexperienced and untrained evaluation manager to formulate competent Terms of Reference and to manage the subsequent evaluation well.

Recommendation [4.3.3] UNICEF evaluation managers should consult generic Terms of Reference when designing an evaluation, should include the topics shown in those generic TORs, and should require the evaluation table of contents to include them, unless there is good reason otherwise. (See Appendices 3 and 4.)

Time Allocated to the Evaluation

Many evaluations we reviewed were conducted in a short time. However, for the same budget, a better result can often be obtained by taking a longer time for the research. Most UNICEF evaluations would benefit from a substantial period (say three months) of preparation before fieldwork. This would allow a thorough literature review, and thorough preparation. For example, all data collection instruments and supporting materials should be translated into the local language and distributed beforehand when one is dealing with a literate population. It would also facilitate correspondence with other agencies interested in the same topic. It also would allow time for a detailed expenditures analysis. In summary, a good part of the evaluation report

(programme profile, costs analysis, review of pertinent research) should be complete before fieldwork.

Many evaluations we reviewed were conducted in a short time. However, for the same budget, a better result can often be obtained by taking a longer time for the research. Most UNICEF evaluations would benefit from a substantial period (say three months) of preparation before fieldwork. This would allow a thorough literature review, and thorough preparation. For example, all data collection instruments and supporting materials should be translated into the local language and distributed beforehand when one is dealing with a literate population. It would also facilitate correspondence with other agencies interested in the same topic. It also would allow time for a detailed expenditures analysis. In summary, a good part of the evaluation report (programme profile, costs analysis, review of pertinent research) should be complete before fieldwork.

Recommendation [4.3.4] Evaluations should not be rushed. Evaluations of alternative intervention strategies generally take longer than performance evaluations and they should be scheduled well ahead of date on which a decision has to be made.

Entrance and Exit Workshops

In a few of the reports we reviewed, we found good use of entry and exit workshops, first to plan the evaluation and get a quick start, and second to provide challenge to the findings, lessons, and recommendations in the final phase of the evaluation, and increase ownership by local organisations. These workshops are an excellent idea, but are not a substitute for a good evaluation report, as they seem to have been in one or two cases.

Recommendation [4.3.5] Each evaluation should include an entry and an exit workshop with key stakeholders.

Language

The evaluation reports were available mainly in English, but also in French, Spanish, and a few other languages. This is fine for internal UNICEF use, but if the work is to be owned locally, some part, perhaps an expanded “executive summary”, must be available in the local language(s).

Recommendation [4.3.6] An expanded executive summary of the evaluation report should be translated into the local language(s). This should be part of the Terms of Reference and appropriately budgeted.

Sign-offs by Specialists

The simplest way to improve quality control of evaluations is to require appropriate sign-offs at the main milestones: evaluation design, mission approval, and draft report. Three types of sign-offs are generally useful to ensuring the quality of an evaluation: management, evaluation specialist and sector specialist. Sign-offs of the evaluation report should be bound into the report.

Checklists should be developed to guide the quality control authorities in regard to the components that must be in place prior to each sign-off. For example, in the evaluation design

phase, a grid that relates the evaluation questions to specific data to be collected should be mandatory. Some possible checklists are shown in Appendices 3 and 4.

Recommendation [4.3.7] To ensure minimum quality standards are met, appropriate management, evaluation, and sector expert sign-offs should be mandatory, particularly at the research design stage of the evaluation.

4.4. Capacity Building

UNICEF country office evaluation focal points are modestly optimistic about the contribution of UNICEF evaluations to building programming and policy capacity among national partners, and building evaluation capacity in UNICEF, but pessimistic about any improvement in evaluation capacity in national partners.

Table 4.4.1: Performance on Capacity Building⁶⁵

Improvement Resulting from UNICEF Evaluations	Programming or Policy Capacity in National Partners	Evaluation Capacity in National Partners	Evaluation Capacity in UNICEF
A great deal	24%	2%	26%
Some, but not significantly changed	64%	82% ⁶⁶	66%
Little	12%	16%	9%

This is not an evaluation about how to improve the evaluation capacity of national partners, and we have not assessed UNICEF’s current approach and activities. However, given the lack of significant impact on the evaluation capacity of national partners that is indicated in Table 4.4.1 above, some of the alternatives that UNICEF might consider include generating (with others) consistent demand for local evaluation consulting, encouraging local/international collaborations (including connections with Evaluation Societies), and training to improve local skills. Requiring international consultants to team with local persons to undertake UNICEF evaluations falls into the second category — encouraging local/international collaborations. Whether it is the best focus for UNICEF efforts, we do not know.

The previous study of the quality of UNICEF evaluations, which looked for evidence of capacity building in the UNICEF evaluation reports themselves in addition to asking country office staff for their opinion, was pessimistic: “*Clear opportunities for building much-needed national capacities were often missed... (but) many evaluations were carried out under circumstances that were not conducive to capacity building. These included desk reviews and quick low-budget studies.*”⁶⁷ They found evidence of significant capacity building in 13% of the evaluation reports, less than the (unverified) numbers UNICEF staff report to us.⁶⁸ (See Table 4.4.1 above.)

⁶⁵ Survey of UNICEF Country Offices, 2002-03, Questions 16, 17, N=44.

⁶⁶ Question 16 category “some but not significant” is summarized here to be comparable with question 17 categories, but in fact has two parts [1] some partners improved evaluation capacity much more than others 39%; and [2] there was some learning generally but not to the extent that capacity changed significantly 43%. See Appendix 2.

⁶⁷ Victora, op. cit., p.12 (D).

⁶⁸ The Victora study team also thought that over-optimism by UNICEF CO staff on the state of capacity building was partly a result of a common misinterpretation of the question: “the open comments revealed that many had

Recommendation [4.4.1] UNICEF should commission an evaluation of alternative ways to improve national evaluation capacity.

4.5. Participation of Children and Women in Evaluation

Most UNICEF CO evaluation focal points think that beneficiaries are adequately consulted and partner/stakeholder organisations appropriately involved. (See Table 4.5.1 below.) There are significant dissents to this view, as exemplified by written comments. For example:

*“The evaluators’ skills in getting genuine children’s participation have been lacking.”*⁶⁹

*“Young people and mothers as rights holders are not adequately involved.”*⁷⁰

*“Mostly our evaluations are for programme people and professionals.”*⁶⁹

*“In most cases partners are involved to a limited extent in designing TORs, selecting consultants and agencies, reviewing and finalizing the final report. They could be more involved in developing the detailed study plan and method, as well as in the field data collection and quality monitoring.”*⁷⁰

Table 4.5.1: Participation in Evaluation⁶⁵

Participation	Beneficiaries Adequately Consulted?	Partner/Stakeholder Organisations Appropriately Involved?
Too little	23%	15%
About right	75%	81%
Too much	2%	4%

Given UNICEF’s rights-based approach, there is a special onus on the organisation to ensure that the rights of children and women are fully respected in UNICEF evaluations. What does that mean? First, and minimally, it requires that those rights be made visible through a clear statement, and that this statement be an integral part of each Terms of Reference for an evaluation.

Second, it requires that the Terms of Reference undergo ethical review to ensure that those rights are not compromised in any part of the research design. (See Section 5.1 below.)

Third, it requires that UNICEF have a system of reporting any adverse events that occur during an evaluation, including any that might be categorised as rights violations. At a minimum, it should be part of the terms of reference of each evaluation that all persons participating in a UNICEF evaluation in any capacity be informed of their right to report any adverse event and where to report it.

misinterpreted the question and referred to capacity built as a consequence of the report recommendations for further training in specific programmatic areas. A specific area mentioned several times was the development of skills that enabled local teams to carry out EPI-type cluster surveys without external assistance.” p.12.

⁶⁹ Appendix 2, Question 11.

⁷⁰ Appendix 2, Question 9.

Finally, it requires that those rights, especially children's right to have their opinions heard, be actively pursued by UNICEF evaluators. We think that something stronger than a Technical Note is needed.

Recommendation [4.5.1] UNICEF Evaluation Office should produce a statement of children's rights as evaluation research subjects and participants.

5. COVERAGE OF ISSUES

5.1. Rights and Equity

A vital question for UNICEF’s performance evaluations to answer is whether the intervention advanced human rights, children’s welfare and equity to an extent commensurate with the investment. For its evaluations of intervention strategies, the issue is what among alternative approaches would have the best rights outcomes? Many evaluations did well on this measure. About one third were rated 4 on 0-4. On the other hand, almost as many, about one in four, were rated at 0 or 1.

Table 5.1.1: Performance on Outcome Analysis⁷¹

“Output Analysis” Quality Criterion	Percent of Reports Acceptable or Better (Score of 2 or more on the 0-4 scale)
Evaluation objectives and questions should reflect UNICEF’s mission and approach to programming, including protection of children’s rights, promotion of their welfare, and gender equality.	70%

UNICEF performance evaluations, on occasion, may legitimately focus on outputs that are only intermediate to rights outcomes. In contrast, evaluations of intervention strategy should always contain a human-rights-based analysis (including a gender analysis). Of all the areas of children’s rights addressed in the evaluations we reviewed, gender analysis was weakest, with the exception of political analysis that was indeed rare.

UNICEF has set out a three-stage methodology for human-rights-based analysis as part of the situation analysis, which we believe, is broadly applicable to evaluations as well: causality analysis, role analysis, and resource analysis, covering behaviours and cultural patterns, norms, politics and legal systems.⁷²

Human-rights-based analysis in evaluations should produce recommendations for direct action on rights where appropriate. It should also cover the topic of improving national capacity for self-assessment and analysis of rights issues (improving capacity to raise awareness and for social mobilisation, and improved planning, technical and professional capacities). Only a few of the evaluations came close to meeting this high standard.

Recommendation [5.1.1] Every UNICEF evaluation of intervention strategy should include a human-rights-based analysis, including identifying systemic problems and potential strategies for addressing them.

Recommendation [5.1.2] Every evaluative human-rights-based analysis should include a gender analysis.

⁷¹ Appendix 1, Criterion 6.

⁷² UNICEF, “Guidelines for Human-Rights-Based Programming Approach”, 1998, CF/EXD/1998-004, B1b.

5.2. Usefulness

Most of the evaluations we reviewed were more like performance evaluations than evaluations of intervention strategy. That is, they focussed mainly on specific performance in unique circumstances that had limited wider relevance. Fifty percent⁷³ of respondents to our survey of country offices reported that evaluations were very useful for assessing and improving the specific initiatives in question, but only 20%⁷⁴ said they were very useful, even in the same country, beyond the specific initiative evaluated.

Our own observations, when we reviewed the evaluation reports, were similar. Limited wider relevance meant that lessons learned were seldom generalised beyond the immediate intervention. Evaluation reports received only 30% of the possible performance score on this criterion.⁷⁵ Wider relevance to UNICEF was seldom discussed. Half of the evaluation reports received a zero on the 0-4 performance scale on this criterion.⁷⁶ However, this is not unacceptable if the reports were performance evaluations, rather than evaluations of intervention strategy. In that case, they might legitimately concentrate on accountability rather than relevance, replicability and scalability. It is difficult to judge because so few of the evaluations were clear about what their goal was — performance assessment or strategy assessment.

Nevertheless, the evaluation reports could have done more to address wider relevance. The degree to which the project, programme or initiative might be replicable in other contexts was seldom described, and reports received only 25% of the possible performance score on this criterion.⁷⁷ This failure is related to the lack of literature review, and a paucity of discussion with other country offices, and other development agencies, during the evaluation design. It may also imply a failure of coordination UNICEF-wide.

The previous study of the quality of UNICEF evaluations reported that about half the evaluation reports reviewed by that study team had “some international relevance”.⁷⁸ This is not to say that the international relevance was well explicated. Other studies of the quality of evaluations by international development agencies have also noted that they seldom grasped the opportunity to draw broad lessons from the specific case.⁷⁹ Given the low level of performance in regard to broad relevance, one wonders whether the consultants had the capability to draw broad lessons.

Recommendation [5.2.1] All UNICEF evaluations of intervention strategy must address the wider relevance of lessons learned, and the replicability and scalability of the successful aspects of the project or programme. (See Appendix 3: Model Table of Contents for an Evaluation of Intervention Strategy.)

⁷³ Appendix 2, Question 12.

⁷⁴ Appendix 2, Question 13.

⁷⁵ Appendix 1, Criterion 26.

⁷⁶ Ibid.

⁷⁷ Appendix 1, Criterion 2.

⁷⁸ Victora, *op. cit.*, p. 27.

⁷⁹ CIDA, *op. cit.*, 1991, p. 26, “Broader lessons learned were not well addressed in the CIDA evaluation studies. Development lessons are not explicitly addressed at all in a third of studies, operational lessons in a quarter, and evaluation lessons in three quarters. Exemplary treatment is relatively rare. This is partly the result of a failure to consider the project in detail within the context of the country programme, to review previous research, and to discuss the matter with other donors.”

5.3. Context and Attribution

Both performance evaluations and evaluations of intervention strategy require careful analysis of external factors to determine what can reasonably be attributed to UNICEF or to a specific treatment/intervention. This is particularly important to evaluation of a rights-based project or programme where understanding the economic, cultural, legal, political, and gender contexts is vital.

Approximately half of the reports we reviewed did well on context analysis and on careful consideration of what reasonably can be attributed to UNICEF. However, about one third did poorly. (See Table 5.3.1 below.)

Table 5.3.1: Performance on Context and Attribution Analysis

Rating	Percent
Zero or 1	30%
2	21%
3 or 4	50%

Source: Assessment of UNICEF Evaluation Reports, Appendix 1, Criterion 4.

Context analysis, leading to scenarios analysis, is also important to determining whether a successful intervention might be replicated in other contexts, and whether it might be expanded to a larger scale successfully.

Finally, the physical environment should receive due attention. Many of UNICEF's interventions, such as clean water projects or the expansion of health or education facilities, have environmental implications. These should be assessed during the evaluation.

Recommendation [5.3.1] Every UNICEF evaluation should include a full profile of the intervention and its context, an attribution analysis where appropriate, and a consideration of replicability, scalability, sustainability, and environmental aspects.

5.4. Analysis of Costs

There is a striking lack of cost consciousness in many UNICEF evaluations. Two in three evaluations did not state the costs incurred by all contributors to the project or programme,⁸⁰ or measure costs or resource use.⁸¹ Fewer, about one third, did not clearly distinguish between the analysis of implementation factors, including costs, and the analysis of results.^{Error! Bookmark not defined.}

⁸⁰ Appendix 1, Criterion 3.

⁸¹ Appendix 1, Criterion 21.

Virtually none compares costs and results. Ninety percent of evaluations received a zero or one on the scale 0-4 in regard to “Costs and results should be directly compared to as full an extent as possible.”⁸² Consequently, it is seldom clear what is to be maximised — the best results for a fixed budget or a target result for the minimum cost.

Table 5.4.1: Performance on Cost Analysis

“Cost Analysis” Quality Criterion	Percent of Reports Acceptable (Score of 2 or more on the 0-4 scale)
State incremental contribution of all stakeholders in cash or kind ^{Error! Bookmark not defined.}	37%
Compare costs and results as fully as possible	10%

Not all consultants are competent cost analysts. If not, then the UNICEF officer in charge of the evaluation needs to fill this gap.

The cost analysis should be largely completed before fieldwork, as part of the profile of the project or programme.

Recommendation [5.4.1] Every evaluation should include a cost analysis.

Recommendation [5.4.2] All performance evaluations should include an efficiency analysis — a comparison of costs and results — and all evaluations of intervention strategy should include a comparative costs analysis for each alternative being assessed.

Recommendation [5.4.3] UNICEF should have guidelines on how to estimate the full cost of a project or programme, including UNICEF staff time, contracts, costs of partners, and costs of participants.

5.5. Analysis of Outputs

Rigorous output analysis tends to be the key missing link in many UNICEF evaluations. The tendency is to describe activities (inputs) and then to over-interpret them in terms of their likely effects on ultimate objectives (outcomes). Careful analysis of the instrumental outputs tends to be neglected. At one end of the causal chain, activities (inputs) are not meaningful, and, at the other end, outcomes are not measurable. This leaves a lot of space for evidence-free speculation and leaps of logic. UNICEF evaluations should pay more attention to the middle ground of “outputs”, which are both meaningful and measurable.

In performance evaluations, output analysis is key to management accountability, and in evaluations of intervention strategy, it is key to assessing replicability and scalability. A recent review by the Asian Development Bank of evaluation capability in borrowing member countries identified the lack of output analysis as the most serious defect.⁸³ This gap is closely linked with

⁸² Appendix 1, Criterion 24.

⁸³ Asian Development Bank, “Special Evaluation Study on Project Performance Management in ADB, Executing Agencies, and Planning Agencies”, August 2002.

the failure of project/programme managers to use output-based contracts in dealing with implementing partners.

About one third of UNICEF evaluation reports almost entirely fail to deal with outputs [a score of 0 or 1 on the 0-4 scale]. In fact, this is a somewhat generous assessment on our part. If the benchmark were the actual measurement, not just mention, of all important outputs, then few evaluations would be highly rated. The paucity of output-based contracts to provide specific and measurable targets for projects and programmes, and the related lack of output analysis in evaluations, are serious deficiencies.

Table 5.5.1: Performance on Output Analysis

'Output Analysis' Quality Criterion	Percent of Reports Acceptable (Score of 2 or more on the 0-4 scale)
Outputs (what is actually produced by the inputs) should be measured ⁸⁴	70%

There is a certain overlap between evaluation and value-for-money audit in regard to outputs. Evaluators could learn from the methods of auditors in this regard, and joint collaborative evaluations/audits are a good approach in some cases.

Recommendation [5.5.1] Every evaluation should include comprehensive measurement of outputs.

Our review of a sample of terms of reference for evaluations indicated that few stated measurable output objectives for the programme or project. We assume that this reflects a lack of output-based implementation contracts/agreements in the programme being evaluated. As the Asian Development Bank recently stated in a study of national evaluation capabilities, accountability and project performance management “necessarily revolves around outputs”. **Error! Bookmark not defined.**

Recommendation [5.5.2] UNICEF should consider making contracts/agreements with implementing agents more output-performance-based, to facilitate results-based management, and to enable good performance evaluation.

5.6. Analysis of Outcomes

Not all evaluations need to measure outcomes. If the project/programme is theory-based, for instance, then outcome analysis may not be necessary. An immunisation programme does not have to re-establish that the vaccine is effective, just that it is administered properly and efficiently to the correct population at the correct time. Similarly, if the project/programme is rights-based, then outcome analysis might not be necessary. It is not necessary to establish each time that gender equality among boys and girls leads to good economic outcomes, for example. On the other hand if the project/programme is institution-based, then it is almost always necessary for the evaluation to measure outcomes. For example, a programme that funds community groups to produce self-help plans in the expectation that income disparity will lessen had better be evaluated by measuring income distribution over time.

⁸⁴ Appendix 1, Criterion 22.

A little fewer than half the evaluation reports received poor marks in regard to outcomes analysis. Past reviews of evaluation quality by CIDA and by USAID have reached similar conclusions.⁸⁵ This is a little difficult to interpret because, as noted above, not all evaluations need to measure outcomes. About one third received exemplary ratings (a score of three or four on the scale 0-4). The performance scores are rather evenly distributed 0 to 3, and 9% received a score of 4.⁸⁶

Table 5.6.1: Performance on Outcome Analysis

'Output Analysis' Quality Criterion	% of Reports Acceptable (Score of 2 or more on the 0-4 scale)
Outcomes/impacts (what the intervention seeks to achieve) should be measured or an appropriate rationale given why not. ⁸⁶	58%

If outcomes are evaluated, then the attribution analysis is important. A reasonable attempt should always be made to estimate how much improvement (or regression) is attributable to UNICEF's inputs and how much to other partners and external factors.

Recommendation [5.6.1] Not every evaluation needs to measure outcomes. In some cases, outcomes can be emphasised too much.

Attribution is often a challenge in evaluation, especially when the programme is a broad one, with many influences and actors. However if outcomes are to be measured, a viable attribution research design is essential.

Recommendation [5.6.2] If outcomes are measured by the evaluators, there should be a well-based attribution analysis as well.

5.7. Analysis of Sustainability

5.8. Analysis of Sustainability

More than half the evaluation reports we reviewed received a zero or one, on the 0-4 scale of quality, for their treatment of sustainability.

⁸⁵ CIDA, op. cit., 1991, p.12. "Purposes achieved" were described by more than 50% of CIDA and USAID evaluations. However, in both cases, about one quarter of the evaluations addressed purposes minimally or not at all.

⁸⁶ Appendix 1, Criterion 23.

Table 5.7.1: Did the evaluation report adequately address the issue of sustainability?

Score	Percent of Reports
Zero or 1	55%
2	20%
3 or 4	25%

This is similar to the quality of CIDA evaluations,⁸⁷ but not as good as USAID evaluations.⁸⁷

Recommendation [5.7.1] If the project, programme, or institution is expected to be sustained without or without continued UNICEF funding then the mechanisms for ensuring this should be explained. If commitments from others are necessary, then these should be described. Self-sustainability of institutions and programmes through cost-recovery should be assessed where relevant.

⁸⁷ CIDA, op. cit., p. 20.

6. REPORTING AND ACTION

6.1. Completeness, Clarity, and Accuracy

Completeness of the evaluations was not easy to judge because the reports varied a great deal in format. The set of topics and sequence were inconsistent, which would make it unnecessarily difficult for managers to read and assimilate the reports.

However, the reports were more clear than complete. Despite gaps, many reports were well written, logical and accessible. This finding is essentially the same as the previous review of the quality of UNICEF's evaluations, which reported that about half the reports were clear and the rest were turgid and verbose.⁸⁸

Table 6.1.1: Are the Reports Clear and Complete?

Score	Clear?	Complete?
Zero or one	17%	29%
2	24%	28%
3 or 4	56%	41%

Less serious than the omission of substantive sections, but still worth some attention, were the flaws in the presentation formats. About a quarter of the reports we reviewed were deficient in their cover page and title, in naming the evaluators, and in having a table of contents. More than half did not contain an executive summary. This was almost exactly the same proportion lacking an executive summary noted by Victora in 1995, in regard to 1992-93 evaluations.⁸⁹

About half the reports appended the terms of reference for the evaluation. However, details of methodology were often missing from the reports, giving the impression of greater interest in findings, lessons and recommendations than in the quality of substantiation. Graphs and other visual presentations of data were relatively rare, and some reports were excessively long. There seems to be a relationship between the quality of the methodology and conciseness of reporting — the better the methodology, the more concise the report.

⁸⁸ Victora, op. cit., p. 14, "Some 40% of the reports provided information that was judged to be good or excellent, 20% fair and 40% poor or none. One particular concern is that 60% of the evaluations failed to include sufficient information on the programme being assessed. 58% of the reports described the problem under study adequately. The objectives were mentioned in almost 90% of the reports but were adequately stated in less than half of these. A common problem was an excessive number of objectives. The reviewers noted that reports lacking precise objectives often suffered from a general loss of direction which in turn affected all subsequent sections."

⁸⁹ Victora, op. cit., p. 14: "Less than half, 44%, of the reports had a summary. However, the degree to which the summary covered different report sections was less often adequate. Summaries often described findings and recommendations, but relatively few included details of context, objectives or, frequently, methodology."

Recommendation [6.1.1] Evaluation reports should be limited to a more or less standard length of presentation, say 50 pages + executive summary + appendices.

Recommendation [6.1.2] Evaluation reports should follow a standard format, unless there is a good reason for varying it.

There are two important completeness issues that could not be assessed from the individual reports⁹⁰: did country offices select the right projects/programmes for evaluation, and were the right issues examined? These issues are outside the scope of our study.

As well, evaluators need to gather information from both those reached and those not reached to be complete. Attrition and non-response should be fully explored, but seldom were. As with many other aspects of quality, there seems to have been little middle ground here. Either the reports were very good or they were poor.

Table 6.1.2: Did the evaluators gather information from those not reached?

Score	Percent of Reports
Zero or 1	66%
2	4%
3 or 4	27%

Accuracy is difficult to judge from the evaluation reports alone because inaccuracy tends to be invisible. For example, if sampling is not done properly or questions are posed in a misleading way, the report may be inaccurate without appearing so. About two thirds of UNICEF country office evaluation staff believe that the reports are generally accurate and about one third think there are important exceptions, where reports are in fact inaccurate.⁹¹

Recommendation [6.1.3] Evaluation reports should discuss data quality and explain the effects of reach constraints, attrition, and/or non-response.

6.2. Quality of Findings, Lessons, and Recommendations

The findings, lessons, and recommendations made in the UNICEF evaluation studies are often high quality. The judgement needs to be qualified by saying that without actually re-doing the evaluation one cannot be sure whether the evaluators were on the right track or not.

Nevertheless, the pattern seems clear. Fifty-seven percent of the reports we reviewed received either a 3 or a 4 on the quality scale 0-4 in respect to recommendations being firmly based.⁹²

This is a good performance. This is similar to findings in the previous study of the quality of

⁹⁰ We did not review the IMEP.

⁹¹ Appendix 2, Question 8.

⁹² Appendix 1, Criterion 27.

UNICEF evaluations, which stated that 59% of studies and evaluations contained useful, relevant recommendations.⁹³

Given the extensive methodological weaknesses in the studies, this relatively good result takes some explaining. If it is true, as it appears to us, that the findings, lessons, and recommendations of the evaluators are often better than their methodologies should produce, then it may be that the consultants are drawing more on their expertise as practitioners in the substantive field of the project, rather than on expertise as evaluators. Of course the ideal is that the evaluation team contain both substantive expertise and evaluation competence.

Alternatively, it may be that the findings, lessons, and recommendations, although they seem good quality to a reader who has nothing else but the report to go on, are in fact more flawed than they appear, and that better quality evaluation methodologies would show this. Plausible reports may have committed all sorts of mistakes by omission or commission that are not visible to the reader who has no other frame of reference than the report itself.

It is also worth noting that, although overall performance was good, a minority of evaluations were very poor in the quality of their findings, lessons and recommendations. Twenty percent received a zero or a one on the 0-4 scale. Again this is similar to the previous study of the quality of UNICEF evaluations, which found that findings were “inappropriate in about one in four reports”.⁹⁴ “The reviewers felt that recommendations could be considerably improved through greater interaction with UNICEF officers.”⁹⁴ In the worst of the evaluations, findings, lessons, and recommendations were little more than “motherhood” exhorting the programme to do better, but exhibiting little grasp of context, strategy, constraints, or resources.

Recommendation [6.2.1] Draft findings, lessons, and recommendations should be subjected to challenge in an “exit workshop”, involving all major stakeholders.

⁹³ Victora, op. cit., p. 11 “In general, the reviewers had a positive impression of the potential usefulness of the reports... In 12 cases (10%) the reviewers specifically felt that the report had no utility at all or even - in a few cases - was misleading.”

⁹⁴ Victora, op. cit., p.16.

6.3. Link to Action

At least three quarters of UNICEF evaluations provide a valid and reliable basis for action. On the other hand, the quality of about one in five is too poor to provide reliable advice and it would be risky to act on their recommendations. Therefore, the first step towards action is for the evaluation manager to decide whether the report is sufficiently well founded, insightful and innovative to provide a basis for action.

The second step is to consider each recommendation and make a decision whether to accept it. Sometimes UNICEF alone can make this decision, but more often, a host-country or other agency is involved. Unfortunately, three quarters of UNICEF country offices in 2002 did not record their decision on each recommendation of their evaluations.

Table 6.3.1: Decisions on Recommendations Recorded?⁹⁵

Response	Recorded	Change in practice since 2000-2001?
Yes	26%	23%
No	74%	77%

Recommendation [6.3.1] The UNICEF country office should record its decision on each evaluation recommendation.

Forty percent of country offices prepare an action plan after each evaluation.

Table 6.3.2: Is an Action Plan prepared?⁹⁶

Response	Action Plan Prepared?	If not Prepared, Other Mechanisms for Follow-up?
Yes	40%	50%
No	60%	50%

Recommendation [6.3.2] For each evaluation plan that contains recommendations that the country office accepts, the evaluation manager should prepare an action plan for approval by the CO senior management.

Recommendation [6.3.3] Where an evaluation recommendation requires action by another agency, UNICEF should ask that agency for a response. If the response is positive, the agency should be asked for an action plan. If an action plan is forthcoming, it should be appended to the evaluation report.

⁹⁵ Appendix 2, Question 20

⁹⁶ Appendix 2, Question 21.

7. FINDINGS

7.1. Overall Quality

The previous review of the quality of UNICEF's 1992-93 evaluations concluded that about 40% were good or excellent, 30% adequate, and 30% poor.⁹⁷ The situation we found a decade later was essentially the same.⁹⁸ The third of evaluations that are poor quality are a serious problem, posing a risk to reputation and of making incorrect decisions based on poor evidence and analysis.

Recommendation [7.1.1] UNICEF needs to improve the quality of the bottom third of evaluations to ameliorate evaluation risk.

A constraint to action to upgrade the quality of the bottom third of evaluations is that, by and large, the CO evaluation focal points are satisfied with quality.

Table 7.1.1: Are the evaluations worth the resources invested?⁹⁹

Response	Percent of Respondents
Yes	93%
No	7%

This question elicited the largest number of optional comments of any question in our survey of the country offices (See Appendix 2). Many used the comments section to qualify the response, generally in the direction of saying that overall value was good, but a significant minority of evaluations did not give value. However, this opinion was not universal. Occasionally an anti-analysis strain showed, as in the comment "Some evaluations ended up destroying office teams. Utilisation was low when they became too academic with insistence on high quality product instead of process..."⁹⁹ There was also a "cheap is good" tone to some comments. For example, "In general UNICEF can get this done more cheaply than say the World Bank so definitely better value than other international agencies."⁹⁹ However, there was also the contrary view: "The bigger the evaluation, the more worthwhile in general."⁹⁹

Recommendation [7.1.2] UNICEF should do fewer evaluations but should achieve adequate quality in all it does do.

7.2. Priorities and Next Steps

The five criteria of quality on which the UNICEF evaluations did best were:

The objectives of the evaluation, and the questions to be answered, should be stated fully and clearly.

⁹⁷ Victora, op. cit., p.19.

⁹⁸ We did not "grade on a curve". If the evaluations had fully met the quality criteria, then they would all have been graded good or excellent.

⁹⁹ Appendix 2, Question 22.

- The evaluation report should be clear, transparent, and easily accessible to the reader. It should be concise, well organised, and logical. The text should be clearly written, supported by tables, figures, and descriptive headings, and led by an executive summary.
- The objectives of the evaluation, and the questions to be answered, should reflect UNICEF's mission and approach to programming, including protection of children's rights, promotion of their welfare, and gender equality.
- Recommendations should be firmly based on evidence and analysis.
- The information gathered, qualitative and quantitative, should in aggregate, be adequate to answer the evaluation questions.

The five criteria of quality on which the UNICEF evaluations did worst were:

- Costs were not well described, and were seldom compared with results.
- The "outputs" of the programme or project were often not adequately described or measured and, with this missing link, the causal chain from activities to outcomes was broken.
- Ethics review was seldom undertaken at the research design stage, and the topic of research ethics was seldom addressed in the reports. It is, of course, vital that the evaluation design be ethical and include ethical safeguards where appropriate, including protection of the dignity, rights and welfare of human subjects, particularly children, and respect for the values of the beneficiary community. We have no opinion on whether there were any ethical problems with the research, such as competence or informed consent, but simply note that the evaluations seldom addressed the topic. The evaluators seldom made a statement about how their objectivity and independence were ensured.
- The evaluations were generally parochial. The degree to which the project, programme or initiative might be replicable in other contexts often was not described.
- Lessons learned often were not generalised beyond the immediate intervention being evaluated to indicate any possible wider relevance to UNICEF.

To improve the quality of its evaluations, UNICEF could focus on risk or excellence, or, of course, both, if sufficient resources can be mobilised.

Poor evaluations pose a risk to UNICEF and its partners. This risk could be lessened by focusing on bringing the worst third up to standard. If the responsibility for evaluation is to remain decentralised, and the quality of the worst third is to improve, then the country offices must be given incentives and tools. The incentives might include positive recognition for good evaluation performance (say, awards for the best ten evaluations each year), and/or penalties for not achieving adequate evaluation standards. Another possible approach that fits the 'minimum standard' option would be to institute a process to certify each country office in regard to evaluation — with a "standards assurance process" for evaluation, perhaps, similar to an ISO quality process.

The tools to upgrade the worst third of evaluations should be appropriately simple. They might include generic frameworks for evaluation terms of reference (including guidelines for processes, timeframes, and budgets), and standard tables of contents for the main types of evaluations (performance evaluations and evaluations of intervention strategy), as recommended by this study. See Appendices 3 and 4. Tools might also include mandatory sign-offs at each major milestone of an evaluation.

Recommendation [7.2.1] UNICEF should ensure that all evaluations meet a minimum standard.

At the other end of the quality spectrum, the problem and the appropriate response are quite different. UNICEF has produced many evaluations that are good, but not excellent. This is perhaps not as much a problem with performance evaluations as with evaluations of intervention strategy. Achieving excellence, starting from this base of good work, would be more difficult and expensive than the “minimum standard” option described above. Three things would probably be key:

- Engaging the best evaluation professionals, which would probably be considerably more expensive than some country offices think they can afford.
- Insisting on rigorous research designs, that are much less impressionistic than UNICEF evaluations have often been.
- Developing methodologies for evaluations of rights-based intervention strategies. The cultural, political, and legal dimensions of rights-based interventions make their evaluation particularly challenging, and UNICEF would probably be breaking new ground in evaluation methodology.

Achieving excellence would require highly trained evaluation managers, as well as highly qualified consultants, and larger budgets to enable more thorough primary data collection and analysis. All this raises issues of how much UNICEF is able and willing to pay for evaluations that are beyond the scope of this study. Perhaps the only way to achieve excellence within an affordable budget is by doing a limited number of evaluations of intervention strategy each year, each involving several country offices. The design and coordination of such evaluations might require leadership from UNICEF headquarters.

Another promising approach is to undertake more evaluations of intervention strategy jointly with other development agencies, including UNDAF partners, the multilateral development banks and the major bilateral agencies. We saw cases where this was done for performance evaluations of jointly funded interventions, and these joint evaluations tended to be better quality than UNICEF-alone efforts. Since there are many areas of rights that are of common interest, it may be possible to institute a series of joint evaluations of intervention strategy with one or more partners among the international agencies.

Recommendation [7.2.2] UNICEF should aim to achieve excellence in some evaluations of intervention strategy each year.

These recommendations relate to UNICEF’s own evaluation capability and performance. They will not necessarily improve the evaluation capability of host-country agencies. Upgrading local capability by involving a few local staff and/or consultants in UNICEF evaluations, often in minor roles, is probably not very effective. Most developing countries need more structural assistance

to improve their evaluation capabilities, such as forming a national evaluation association with affiliation with an international professional association. All to these alternative strategies need to be fully assessed, and choices made by UNICEF in regard to its highest priorities for improving evaluations.

Recommendation [7.2.3] UNICEF should formulate a medium-term “evaluation quality improvement plan” as the action plan arising from this study.

7.3. List of Recommendations

The following is a list of recommendations in the order in which they appear in this report. For a summary and grouping of recommendations, see the Executive Summary.

NYHQ Evaluation Office and Regional Offices should develop a three-year evaluation plan in conjunction with the country offices to coordinate evaluation activity and ensure adequate coverage of key issues. The plan should be updated annually.

UNICEF needs some means of frequently updating country office awareness of what evaluations are being started and completed in other country offices.

The full text of completed evaluations from the current year and two previous years should be accessible on the Internet in Adobe format, and key worded for easy search and access.¹⁰

Every UNICEF evaluation should be written for a broad audience within the organisation, giving sufficient background and context to a reader unfamiliar with the country and programme, and explicitly describing the international relevance of points made in the evaluation.

To facilitate openness and provide incentive to improve quality, all UNICEF evaluations that cost more than US\$10,000 should be available to the public on the open website.

UNICEF should require that an electronic copy of all Terms of Reference should be submitted to the Regional Office and NYHQ before an evaluation is contracted.

UNICEF Terms of Reference should include a model Table of Contents for the evaluation report, such as those shown in Appendices 3 and 4.

UNICEF should develop a standard electronic form, with guidelines, for Terms of Reference for evaluations.

Substantial evaluations, say those budgeted at more than US\$25,000, should begin with an explicit research design exercise to produce an “evaluation framework”, and this framework should be referenced by the TORs for the subsequent evaluation.

The Terms of Reference for each evaluation study should state whether the study is to contain a situation analysis, an evaluation of intervention strategy, and/or a performance evaluation, and for each of these that it does contain, an appropriate methodology should be described.

Each evaluation study should have a methodological review and challenge by a peer in the country office, or, when the evaluation is important and difficult, by the Regional Office, or, when more appropriate, by someone in another CO. For each evaluation budgeted at over US\$10,000, this review should be based on a full framework.⁷

NYHQ should retain a sampling expert and make his or her time available to any CO designing an evaluation, via the Internet.

UNICEF should schedule evaluations of intervention strategies and performance evaluations at points where most can be learned, and/or when decisions depend on the evaluation (not only when funding decisions are imminent).

UNICEF should state an evaluation research ethics review policy, including a policy on subject competence and informed consent, and a policy on adverse event reporting.

UNICEF should institute a research ethics board, and the board should apply for accreditation, perhaps by the AAHRPP. **Error! Bookmark not defined.**

Each evaluation report should contain a statement of how the objectivity and independence of the evaluators and their report was ensured.

Few evaluations of intervention strategy should take less than seven months from contracting to completion, and few performance evaluations less than four or five months.

Each evaluation should include an entry and an exit workshop with stakeholders.

An expanded executive summary of the evaluation report should be translated into the local language(s). This should be part of the Terms of Reference and appropriately budgeted.

To ensure that minimum quality standards are met, appropriate management, evaluation and sector expert sign-offs should be mandatory at each of the three main milestones of each evaluation. Formal sign-offs (on a single standard-format page) should be bound in the final report.

UNICEF should commission an evaluation of alternative ways to improve national evaluation capacity.

UNICEF should produce a statement of children's rights as evaluation research subjects and participants.

Every UNICEF evaluation of intervention strategy should include a rights analysis, including identifying systemic problems and potential strategies for addressing them.

Every evaluative "rights analysis" should include a gender analysis.

All UNICEF evaluations of intervention strategy must address the wider relevance of lessons learned, and the replicability and scalability of the successful aspects of the project or programme. (See Appendix 3: Model Table of Contents for an Evaluation of an Intervention Strategy.)

Every UNICEF evaluation should include a full profile of the intervention and its context, an attribution analysis where appropriate, and a consideration of replicability, scalability, sustainability, and environmental aspects.

Every evaluation should include a cost analysis.

All performance evaluations should include an efficiency analysis — a comparison of costs and results — and all evaluations of intervention strategies should include a comparative cost analysis for each alternative being assessed.

UNICEF should have guidelines on how to estimate the full cost of a project or programme, including UNICEF staff time, contracts, costs of partners, and costs of participants.

Every evaluation should include comprehensive measurement of outputs.

UNICEF contracts/agreements with implementing agents should be output-performance-based. This is essential to a results-based approach to good management, and to enabling good performance evaluation.

Not every evaluation needs to measure outcomes. In some cases, outcomes can be emphasised too much.

If outcomes are measured by the evaluators, there should be a well-based attribution analysis as well.

If the project or programme is expected to be sustained, without or without continued UNICEF funding, then the mechanisms for ensuring this should be explained. If commitments from others are necessary, then these should be described. Self-sustainability through cost-recovery should be assessed.

Evaluation reports should be limited to a more-or-less standard length of presentation, say 50 pages + executive summary + appendices.

Evaluation reports should follow a standard format unless there is a good reason for varying it.

Evaluation reports should discuss data quality, and explain the effects of reach constraints, attrition and/or non-response.

Draft findings, lessons and recommendations should be subjected to challenge in an “exit workshop”, involving all major stakeholders.

The UNICEF country office should record its decision on each evaluation recommendation: either reject it as unsubstantiated, as wrong in substance, or accept it.

For each evaluation plan that contains recommendations that the country office accepts, the evaluation manager should prepare an action plan for approval by the CO senior management.

The approved action plan should become an appendix to the evaluation report.

Where an evaluation recommendation requires action by another agency, UNICEF should ask that agency for a response. If the response is positive, the agency should be asked for an action plan. If an action plan is forthcoming, it should be appended to the evaluation report.

UNICEF's highest priority in evaluation should be improving the quality of the bottom third of evaluations. Its evaluation management should be targeted mainly to this goal in the short term.

UNICEF should do fewer evaluations, but should achieve adequate quality in all it does do.

UNICEF should ensure that all evaluations meet a minimum standard.

UNICEF should achieve excellence in some evaluations of intervention strategies each year.

UNICEF should formulate a medium-term "evaluation quality improvement plan" (equip), as the action plan(s) arising from this study.

APPENDICES

Appendix 1: Evaluation Report Quality	47
Appendix 2: Country Office Feedback on Evaluation Quality	57
Appendix 3: Model Table of Contents for an Evaluation of Strategy	85
Appendix 4: Model Table of Contents for a Performance Evaluation	89
Appendix 5: Terms of Reference	91
Appendix 6: UNICEF Response to Specific Recommendations	97

APPENDIX 1: EVALUATION REPORT QUALITY

General Comments on Quality:

Cover page and title	yes = 75%
Date	yes = 96%
Author named	yes = 77%
Table of contents	yes = 75%
Executive summary	yes = 57%
TORs attached?	yes = 45%
Number of pages	

At what level is the evaluation focussed?

[1] The whole UNICEF country programme	10%
[2] A major programme	60%
[3] A project or activity, as part of a programme	30%

Criterion 1: The project, programme or initiative to be evaluated should be clearly described, including the logic of its links between activities, outputs, and outcomes.

Score:

Total Score	170
% of Possible Score	57%
% Zeros	4%
% Score = 1	28%
% Score = 2	25%
% Score = 3	23%
% Score = 4	20%

Criterion 2: The degree to which the project, programme or initiative might be replicable in other contexts should be described.

Score:

Total Score	76
% of Possible Score	26%
% Zeros	57%
% Score = 1	12%
% Score = 2	12%
% Score = 3	9%
% Score = 4	9%

Criterion 3: The incremental contributions of UNICEF and other stakeholders, in cash or in kind, should be stated.

Score:

Total Score	103
% of Possible Score	34%
% Zeros	39%
% Score = 1	24%
% Score = 2	8%
% Score = 3	20%
% Score = 4	9%

Criterion 4: The context should be analysed in sufficient detail to identify external factors contributing to success or failure, and this analysis should be used to support informed judgements about what results may reasonably be attributed to UNICEF's intervention, and what to other factors.

Score:

Total Score	162
% of Possible Score	54%
% Zeros	19%
% Score = 1	11%
% Score = 2	21%
% Score = 3	35%
% Score = 4	15%

Criterion 5: The objectives of the evaluation, and the questions to be answered, should be stated fully and clearly.

Score:

Total Score	200
% of Possible Score	70%
% Zeros	4%
% Score = 1	19%
% Score = 2	11%
% Score = 3	29%
% Score = 4	32%

Criterion 6: The objectives of the evaluation, and the questions to be answered, should reflect UNICEF's mission and approach to programming, including protection of children's rights, promotion of their welfare, and gender equality.

Score:

Total Score	185
% of Possible Score	63%
% Zeros	16%
% Score = 1	9%
% Score = 2	17%

% Score = 3	21%
% Score = 4	33%

Criterion 7: The evaluation should use standard UNICEF evaluation criteria.

Relevance	Analysed?	yes = 67%
Efficiency	Analysed?	yes = 32%
Effectiveness	Analysed?	yes = 85%
Impact	Analysed?	yes = 57%
Sustainability	Analysed?	yes = 43%

Score:

Total Score	149
% of Possible Score	50%
% Zeros	7%
% Score = 1	25%
% Score = 2	36%
% Score = 3	24%
% Score = 4	7%

Criterion 8: The evaluation report adequately addresses the issue of sustainability.

Score:

Total Score	105
% of Possible Score	35%
% Zeros	46%
% Score = 1	8%
% Score = 2	20%
% Score = 3	12%
% Score = 4	13%

Criterion 9. The evaluation report adequately addresses the issue of relevance.

Score:

Total Score	150
% of Possible Score	51%
% Zeros	22%
% Score = 1	12%
% Score = 2	25%
% Score = 3	20%
% Score = 4	20%

Criterion 10: For evaluations of humanitarian response, in particular at the sector or whole country level, the evaluation should address coverage, coordination, coherence, and protection.

Score:

Total Score	22	
% of Possible Score		55%
% Zeros		20%
% Score = 1		0%
% Score = 2		5%
% Score = 3		3%
% Score = 4		3%

Criterion 11: The methodology of the evaluation should be practical and appropriate to the questions posed, while providing a complete and fair assessment.

Score:

Total Score	174	
% of Possible Score		58%
% Zeros		11%
% Score = 1		15%
% Score = 2		24%
% Score = 3		33%
% Score = 4		17%

Criterion 12: The evaluation design should be ethical and include ethical safeguards where appropriate, including protection of the dignity, rights, and welfare of human subjects, particularly children, and respect for the values of the beneficiary community.

Score:

Total Score	50	
% of Possible Score		15%
% Zeros		67%
% Score = 1		3%
% Score = 2		5%
% Score = 3		11%
% Score = 4		19%

Criterion 13: The evaluators should make a statement about how objectivity and independence were ensured.

Score:

Total Score	85	
% of Possible Score		29%
% Zeros		54%
% Score = 1		7%
% Score = 2		15%

% Score = 3	19%
% Score = 4	5%

Criterion 14: The constraints of the evaluation, and the perspective from which the intervention is evaluated, should be clear so the reader can assess the validity of the evaluators' judgements.

Score:

Total Score	161
% of Possible Score	54%
% Zeros	23%
% Score = 1	13%
% Score = 2	9%
% Score = 3	36%
% Score = 4	19%

Criterion 15: The evaluation report should describe the kinds of UNICEF and other key stakeholders' follow-up action expected in regard to lessons and recommendations.

Score:

Total Score	176
% of Possible Score	59%
% Zeros	13%
% Score = 1	17%
% Score = 2	16%
% Score = 3	28%
% Score = 4	25%

Criterion 16: The information gathered, qualitative and quantitative, should, in aggregate, be adequate to answer the evaluation questions.

Score:

Total Score	181
% of Possible Score	61%
% Zeros	9%
% Score = 1	15%
% Score = 2	21%
% Score = 3	33%
% Score = 4	20%

Criterion 17: Sampling is used when appropriate and is designed to produce representative data.

Score:

Total Score	100	
% of Possible Score		39%
% Zeros		38%
% Score = 1		11%
% Score = 2		13%
% Score = 3		21%
% Score = 4		8%

Criterion 18: The data collection instruments should be able to provide the measurements needed.

Score:

Total Score	157	
% of Possible Score		52%
% Zeros		17%
% Score = 1		15%
% Score = 2		20%
% Score = 3		37%
% Score = 4		11%

Criterion 19: Information should be gathered from eligible persons not reached in addition to those reached.

Score:

Total Score	88	
% of Possible Score		31%
% Zeros		57%
% Score = 1		9%
% Score = 2		4%
% Score = 3		11%
% Score = 4		16%

Criterion 20: The report should distinguish clearly between implementation factors (inputs, process milestones, activities) and results (outputs and outcomes/impacts).

Score:

Total Score	145	
% of Possible Score		49%
% Zeros		19%
% Score = 1		13%
% Score = 2		35%
% Score = 3		23%
% Score = 4		9%

Criterion 21: Inputs (costs/resources) should be measured.

Score:

Total Score	92	
% of Possible Score		31%
% Zeros		43%
% Score = 1		20%
% Score = 2		16%
% Score = 3		15%
% Score = 4		7%

Criterion 22: Outputs (what is actually produced by the inputs) should be measured.

Score:

Total Score	145	
% of Possible Score		48%
% Zeros		24%
% Score = 1		9%
% Score = 2		31%
% Score = 3		21%
% Score = 4		21%

Criterion 23: Outcomes/impacts (what the intervention seeks to achieve) should be measured or an appropriate rationale given why not.

Score:

Total Score	135	
% of Possible Score		46%
% Zeros		23%
% Score = 1		19%
% Score = 2		24%
% Score = 3		24%
% Score = 4		9%

Criterion 24: Costs and results should be directly compared to as full an extent as possible.

Score:

Total Score	31	
% of Possible Score		11%
% Zeros		74%
% Score = 1		15%
% Score = 2		9%
% Score = 3		1%
% Score = 4		0%

Criterion 25: Accountability for results (who is accountable for what, and their performance) should be appropriately analysed and reported.

Score:

Total Score	118	
% of Possible Score		39%
% Zeros		31%
% Score = 1		19%
% Score = 2		20%
% Score = 3		24%
% Score = 4		7%

Criterion 26: Lessons learned should be generalised beyond the immediate intervention being evaluated to indicate any possible wider relevance to UNICEF.

Score:

Total Score	91	
% of Possible Score		30%
% Zeros		48%
% Score = 1		12%
% Score = 2		20%
% Score = 3		11%
% Score = 4		9%

Criterion 27: Recommendations should be firmly based on evidence and analysis.

Score:

Total Score	185	
% of Possible Score		62%
% Zeros		8%
% Score = 1		12%
% Score = 2		23%
% Score = 3		40%
% Score = 4		17%

Criterion 28: Priorities for action should be made clear.

Score:

Total Score	163	
% of Possible Score		56%
% Zeros		18%
% Score = 1		11%
% Score = 2		24%
% Score = 3		23%
% Score = 4		23%

Criterion 29: The evaluation report should be complete, with all of the necessary elements of a good report present.

Score:

Total Score	166	
% of Possible Score		56%
% Zeros		4%
% Score = 1		25%
% Score = 2		28%
% Score = 3		29%
% Score = 4		12%

Criterion 30: The evaluation report should be clear, transparent, and easily accessible to the reader. It should be concise, well-organised, and logical. The text should be clearly written, supported by tables, figures, and descriptive headings, and led by an executive summary.

Score:

Total Score	191	
% of Possible Score		65%
% Zeros		1%
% Score = 1		16%
% Score = 2		24%
% Score = 3		35%
% Score = 4		21%

APPENDIX 2: COUNTRY OFFICE FEEDBACK ON EVALUATION QUALITY

Questions and Responses:

1. Were there one or more evaluations undertaken for your office during 2000 and 2001 that were not submitted to New York headquarters?

No	62%
Yes	38%
Number of responses:	45

If there were more evaluations done than submitted, how many were there approximately?

In aggregate, the sample of country offices reported that they had completed 74 evaluations during 2000 and 2001 that had not been submitted to NY HQ — approximately 37 per year, or, extrapolating from the sample, about twice that for all COs.

2. In your opinion, was the overall level of evaluation activity supported by your CO in 2000 and 2001:

Too little	25%
About right	75%
Too much	0
Number of responses:	47

Selected Comments:

As many small-scale evaluations were conducted, it was not always possible to give attention to the quality of those studies.

The limited human and financial resources of Small Islands Developing States (SIDS) influence strongly the ability to set-up and maintain systems for social data gathering, processing and analysis, monitoring and evaluation. In general there is a weak appreciation of the value of monitoring and evaluation in the region. More evaluation activities are needed to improve programme and project planning and policy development.

Car nous avons donné priorité à l'enquête à Indicateurs multiple des Objectifs de la Décennie qui nous a donné plusieurs informations par l'évaluation du Programme Pays prévue.

In 2000 we have managed about right number of evaluations since we have had MTR that year. In 2001 the overall level of evaluation activity was too little.

A choice was made to limit the number of evaluations and research activities to only those that would inform programme development.

Too little compared to the activities planned (3 evaluations completed vs. 9 evaluations planned). But 2001 was the first year of the country programme.

2000 was last year of country programme (a lot of evaluation), 2001 was first year (a lot of baseline surveys, little evaluation).

The Country Office has only started implementing its 1999-2003 Programme in 2000. In 1999 evaluation activities were kept on hold due to an audit and its aftermath. So, most evaluations were planned for the Mid Term or end of programme cycle reviews.

Many studies and evaluations need to be undertaken and are now planned for 2003.

There were several evaluations/studies undertaken by our ---- offices but the overall quality of these is poor. In one year there is one or two worth wide sharing.

In general our office pays more attention to studies than to evaluation of programme interventions.

It is very difficult to find an appropriate person in our country who can do evaluation properly; therefore we have to rely on international consultants – needs capacity building on evaluation.

Three evaluations were completed in 2000. Two of them were for Education and the rest is for Integrated Basic Services. There was no evaluation done in 2001. We should have thematic evaluations as well.

We undertook a reasonable amount of evaluation I think for a medium-sized office of 20 –30 people, which can do a maximum of three major external evaluations per year.

3. Were there particular areas or topics that UNICEF, globally or in your country, has not evaluated sufficiently?

No	25%
Yes	75%
Number of responses:	40

If some areas or topics were not evaluated sufficiently, what were they?

Children's participation was not sufficient. However, this has been integrated into this year's activities. The timing before now was not right.

Le trafic des enfants, les enfants de la rue, les enfants sans support qui ont abandonné.

Quality of education (especially as it relates to female empowerment), violence and children, and budgetary allocation for children.

Violence against children and women is not evaluated sufficiently.

Training and social mobilization were not evaluated sufficiently.

Child Labour, HIV/AIDS, Juvenile Justice were not evaluated sufficiently.

Support to psychosocial interventions in emergency situations; education of behavioural change types of interventions; learning achievements; child labour prevention strategies; were not evaluated sufficiently.

I think most long-term interventions are not properly and adequately evaluated.

The effectiveness, efficiency and the long-term impact of emergency operations were not evaluated sufficiently.

As at 2001, the CO had not conducted an overall evaluation of the --- programme for the last three country programmes (i.e. for over 7 years). As such, in late 2001 it was determined that it is important for the CO to embark on and evaluation of the Country Programme (CP) at least covering the last two programme cycles. This evaluation was conducted in 2002 with a two-fold purpose. First, it was necessary to closely examine the programme to determine where UNICEF was correctly addressing needs, and which aspects of the programme were strong and weak, in order to analyse how best to respond to the various communities within ----- . Second, the evaluation would serve as a critical tool in the formation of UNICEF's new Country Programme, which will begin in 2004.

The impact of civil war on the children, and the humanitarian assistance/aid vis-à-vis peace were not evaluated sufficiently.

Cost effectiveness is almost never evaluated adequately.

Child trafficking, and children without primary caregivers, was not evaluated sufficiently.

Social policies, decentralisation, participation, child protection, and adolescents and youth were not evaluated sufficiently.

HIV/AIDS, and ECD were not evaluated sufficiently.

The Country programme was not evaluated sufficiently.

Integral evaluation to incorporate the cost/benefit relation of the interventions carried out by UNICEF during the 2000-2001 period at state level, where the Technical Cooperation agreements were carried out.

Area focused/based programmes were not evaluated sufficiently.

The effectiveness of Country Programme, Effectiveness of Partnership with Government was not evaluated sufficiently.

For Evaluation in Health sector, we are planning to have 2-3 evaluations in 2003. Evaluation on Programme Communication is on going. We should have an evaluation in Child Protection and Communication (public relations).

Training provision and Impact of Parenting Skills Education Projects were not evaluated sufficiently.

Child Protection – vulnerability needs and services. Child Protection – children exposed to high risk.

Project interventions in the areas of protection, gender, disaster preparedness, as well as some programme strategies and approaches common to all projects (inter-sectorality, working with NGOs).

HIV/AIDS, youth, crime and violence.

We have had sufficient number of evaluations, but didn't get the expected outcomes. E.G. – bad timing; right content in the reports, meaning findings and recommendations due to either poorly organised evaluation' process, or lack of consultant' capacity / e.g. knowledge of local situation.

Immunization coverage needs evaluation.

4. How accessible do you find evaluations undertaken by other COs?

Highly accessible	17%
Only accessible with some difficulty	77%
Inaccessible	6%
Number of responses:	48

Selected Comments:

The evaluation reports could be accessed on the UNICEF Intranet, but the web site is not updated regularly.

If an evaluation concerns a specific situation that is already well known and understood, the evaluation is accessible. If not, the report is often difficult to relate to.

It is often difficult to get electronic copies of evaluations from other countries.

Generally accessible through e-mails, UNICEF intranet, ordinary distribution by country offices.

Inaccessible. Moreover, without knowing which is a good evaluation, I have little incentive to look.

One needs to look into the evaluation database, identify the evaluations on the topic of your interest and ask the officer in that country office to send you a copy if you want to have access to the entire document. Evaluations will be accessible if you make some effort.

The main source to access evaluations is the Intranet. Only recently (not all) evaluations completed in 2000 and 2001 were posted there. Access to Intranet is not always successful. Furthermore downloading of some existing documents is not easy or they are not accessible (e.g. some evaluations posted recently under Education theme). In addition evaluations are posted together with studies and surveys, which makes the searching process longer. Circulation of high-quality evaluations to the M&E focal points by the Evaluation Office might ease the process of keeping them abreast. The selection itself would help M&E officers/focal points to understand the translation of criteria of the Evaluation Office for high-quality evaluations into practice.

Only accessible with difficulty. Can only speak of my experience over the past 6 months that I have worked in UNICEF. My rating is based on a previous experience when I needed some information from the regional office and it was accessible but with some difficulty.

Within the TACRO region, Spanish fluency is a must and we do not possess this.

The Evaluation CD ROM of summaries was a good initiative. However, the summaries do not identify the good evaluations from the weak ones and do not provide a quick guide. As a result, the process of accessing the evaluation reports of other countries was long. But now with the Evaluation site on Internet, it became much easier to access the evaluations summaries of other countries

Evaluations are not available at UNICEF's database; although theoretically speaking they should be there.

I wrote to Evaluation Office/HQ recently asking for one ("Training as a strategy for capacity building: Taking stock in Myanmar.") and I am still waiting for a response. To be perfectly honest, if the office invested in searching the evaluation databases previously provided by the HQ, it may not have been such a problem, but we don't think of going back to this database

Usually we request by email, but occasionally it takes a long time to obtain the study.

Communication between COs and sharing of information is good; Cluster networks are good channels for sharing information; Evaluation Data-base is also accessible to everyone.

If the report is posted in the web is very accessible if not then there are some difficulties.

Very seldom does the CO receive evaluations from other COs.

Difficulties to access the Intranet (deficiency of local internet connection).

Unless there is systematic sharing between CO's via RO, it is often through personal contacts only that we know of evaluations.

If we request to other COs by email specifically, it is accessible. However, I am not aware of the mechanism of sharing evaluations recently undertaken by other COs. Evaluation database in UNICEF intranet could be improved in terms of retrieving the document and its user friendliness.

In most cases only the executive summary is available.

Some evaluations are not displayed on intranet because considered relevant only for the CO, e.g. evaluation of IE&C campaigns.

Information on what is being done in other countries is not widely shared.

I have never asked another country office to send me an evaluation that I found on the evaluation database. However, during the AfreA conference (the African Evaluation Network) held in Nairobi in June 2002, many evaluations of good quality on various topics were presented, but a UNICEF office did not fund these.

Accessibility not very encouraging or conducive. I cannot judge quality because I have not read any up to now.

Using UNICEF Web page it is very easy to find/search for various reports from different countries.

We were able to access some studies related to child abuse via the Intranet

I usually get them via personal recommendation, as most of the time evaluations done in the region are more relevant to what we do here.

When was the last time you read an evaluation study from a country office other than your own?

In the past month	18%
In the past quarter	47%
In the past year	20%
Very seldom or never	15%
Number of responses:	40

Selected Comments:

Evaluations from other offices would be read more frequently if they were more accessible. Establishment of a searchable data base or a periodical with abstracts of studies in other COs will improve use.

It is possible to find some studies on the Intranet (though often you find that only the summary is available). More often, project officers who know about projects/evaluation in other COs send requests directly to those COs to receive a copy of evaluation report from them, etc.

The only evaluation report I read I got on the cluster meeting, not on the Intranet.

I can't remember the last time I saw an evaluation study from another country office cross my desk! As I was writing this comment a special issue of UNICEF and UNAIDS Evaluations of HIV/AIDS Programmes in sub-Saharan Africa came through the pouch, and will be skimmed, at a minimum.

5. In regard to the evaluations supported by your CO in 2000 and 2001, were they based on high quality Terms of Reference?

Almost always	41%
Generally, but with some important exceptions	46%
Often not	13%
Number of responses:	46

Selected Comments:

Very often Terms of Reference were shared and commented upon by other UN agencies.

The terms of reference for this end decade NPA progress evaluation were of high quality, but this cannot prejudice the lack of updated and reliable data met at the level of the various ministries concerned.

Some of the TORs are rather brief and do not contain all details. In addition there are missing TORs from the period under reporting

Indeed all evaluations supported by the UNICEF country Office are purely based on high quality terms of reference. The practice is that TORs are developed and circulated to the country team for observations and comments. This process is continuous until the team is satisfied and the TORs are adopted.

We generally take a lot of effort in developing the TORs

Large scale evaluations with significant funding and carried out by professionals are based on high quality Terms of Reference. However, majority of small-scale evaluations often have weak TORs.

POs don't always consult the M&E team.

TOR are continuously improving in our office, but the problem is usually in the quality of the evaluation itself.

Since TORs are not always shared by the programme section with the ME section, unable to comment on accurately.

There is no practice of sharing TORs outside CO for feedback; no feedback has ever been received; Staff have no way of assessing the quality level of the TORs.

Some TORs are considered well prepared. Needs special technical guidance.

We have recently introduced a more rigorous system of reviewing and approving TORs. This has resulted in some improvements in the quality of TORs.

We put a lot of time and effort into the planning phase and joint work with counterparts on the TORs.

There is no mechanism to review TOR for each study before it is conducted. Usually TORs are circulated only within the respective section itself, and go to SPC for comments/approval. Few Programme Officers decide to circulate draft TOR to staff outside the section for comments, partly because they have no time to do so, or could be because they do not expect to receive many useful comments (which may not be true). The few that are circulated to all staff are generally of good quality, and further become improved after incorporating the comments from others.

Inter-sectoral co-ordination and co-operation can be improved.

Special attention was paid on high quality TORs. All TORs enclosed to evaluation reports were submitted to NYHQs.

Souvent les Termes de référence ne précisent pas bien les objectifs de l'évaluation, qui peut ou qui utilisera les résultats de l'évaluation et comment ça va se faire c'est à dire les méthodes de collecte des données.

Very often Terms of reference were shared and commented upon by other UN agencies (see health programme evaluations: - vaccination coverage in the ----).

6. Were these evaluations timely in their findings and recommendations?

Almost always	47%
Sometimes	53%
Seldom or never	0
Number of responses:	45

Selected Comments:

Most of the evaluations are initiated in a timely manner. Sometimes the final analyses are not completed in the time frame for the findings to be incorporated during the designing of the next phase.

Generally, it took longer to produce findings than envisaged.

Quelquefois les recommandations ne découlent pas des conclusions et nous avons souvent du mal à identifier à qui s'adresse les recommandations, les recommandations souvent vagues, peu réalistes et ne sont pas classées par ordre de priorités.

Evaluations generally took longer to produce findings than envisaged.

An end-decade evaluation was undertaken with the assistance of a consultant working with concerned ministries, in order to meet the deadline set by NYHQ.

The evaluation reports that I have had the opportunity of reading have been relevant and timely in their findings and recommendations. This particular evaluation --- was both relevant and timely for the new CP. It should be noted also that there have been instances where evaluation reports have been delayed.

The presentation of the final documents took in all instances longer than expected. However, in most cases, the preliminary results were disseminated immediately after fieldwork.

At times they are commissioned a little late, and with some delays in the implementation, and we end up having the results discussed or available when it is too late.

In our experience, evaluations are most frequently conducted at end of project, and are not formative, but summative,

Sometimes not meeting the deadlines. This is due the procedures of endorsing/approval the results.

Some issues were sensitive to the government, and so were not always reported

The results sufficiently timely to feed into programme development.

7. In your opinion, were the evaluation findings generally accurate?

Generally accurate	63%
Generally with some important exceptions	38%
Too often inaccurate	0
Number of responses:	48

Selected Comments:

UNICEF evaluations generally give accurate analysis on quantifiable indicators. But in most cases, in-depth analysis beyond the quantitative aspect is lacking. For example, some of the survey questions do not allow for accurate or in-depth analysis (e.g. a question given to head teachers of IDEAL project primary schools: "Do teachers in your school use the Multiple Ways of Teaching and Learning method in the classroom?" Choices of answers: "Some do" "Few do" "None of them do". The report concludes by saying "A considerable number of teachers use MWTL method"). Furthermore, even when the questionnaire is formulated in a proper way, not much effort is given in analyzing the data collected.

The evaluation findings were generally accurate as the country has not made significant progress towards achieving set goals and the major constraint was the lack of updated data.

Most of the time we find the evaluation findings are accurate.

For the 1st evaluation, findings were generally accurate; for the 2nd there were some question marks.

The issue is not just the accuracy of the findings but also the political feasibility of acceptance by stakeholders and subsequent utilization.

8. Were partner/stakeholder organisations appropriately involved in the various phases of evaluations?

Too little	15%
About right	81%
Too much	4%
Number of responses:	47

Selected Comments:

In most cases, partners are involved to a limited extent in designing the TOR, selecting consultants/agencies, reviewing and finalising the report. They could be more involved in developing the detailed study plan and method, as well as in the field data collection and quality monitoring.

Ownership is very crucial for the quality of the evaluation and the implementation of follow-up actions.

All evaluations were participatory, with stakeholders involved from the phase 1 – drafting a TORs.

Young people and mothers as rights holders were not adequately involved in the elaboration of TORs, the process of evaluation and follow up.

For some major evaluations (such as micro credit), a major effort made to involve stakeholders.

The consultant has undertaken the evaluation in collaboration with the various ministries concerned.

Evaluation are usually done with partners' full cooperation and involvement.

The participatory evaluations sometimes influenced by the project staff and some of the consultants may seek to satisfy the organisation for future assignment.

The stakeholders usually play the lead role with only technical advice coming from the CO.

For some studies shared with ME section, mentioning on this aspect in their methodology sections is too little to make a valid comment on.

Generally one government organization (Ministry of Health) dominates the whole process.

The Ministry of Health (MoH) took the leading role in this evaluation. UNICEF role was limited to co-financing and some technical advice. Other partners in this project include the Ministry of Municipalities, Salt producers and the households. The Municipalities participated in this evaluation, however, households were not involved directly neither were the salt producers.

New evaluation culture here so we need to brief stakeholders carefully.

9. Did ethical problems, such as protecting human subjects or respecting community values, arise during the evaluations?

No	65%
Yes, but were almost always fully resolved	26%
Yes, and were occasionally unresolved	7%
Yes, and were frequently unresolved	2%
Number of responses:	43

Selected Comments:

We have had two significant cases of ethical problems and important lessons were also drawn from these.

The question is not quite clear but the ethical; question of sanctions were raised by The Situation of Children in Iraq, an assessment based on the United Nations Convention on the Rights of the Child.

In areas of war zones there is conflicting interest among the factions, tribal and government groups (vested interest).

Don't know!!!

This aspect needs to be better understood by researchers and UNICEF staff.

No, because all evaluation, except internal ones, need to get the consensus of the line ministries and centralized management).

Not paid sufficient up front attention. Research ethics is not a regular issue to be considered before undertaking evaluations.

Special attention should be paid to a briefing and choice of background materials when hiring an international consultant. From our experience the best choice is to have an external evaluation team consists of international and national experts.

10. Were intended beneficiaries adequately consulted in the various phases of the evaluations?

Too little	23%
About right	75%
Too much	2%
Number of responses:	44

Selected Comments:

They are consulted, but usually to a limited scale in a very fixed style. Only few studies have properly utilized participatory approaches.

Consulted yes, but getting feedback needs quite often additional efforts and facilitation.

The participation of beneficiaries should be part of the TOR (and again to underline the importance of high quality of TOR).

The evaluation has been undertaken mainly with the ministries concerned.

We feel there is room for improvement in this area.

Through qualitative assessment methods (like Focus Groups Discussions, etc).

Intended beneficiaries were adequately consulted in case of community-based programmes. But sometimes the evaluation falls short to reflect the opinion of the stakeholders.

The participatory approach to evaluation where beneficiaries involved in the various phases was not fully introduced to the researchers

This has posed difficulties. Major beneficiaries of UNICEF are children. The evaluator's skills in getting genuine children's participation has been lacking.

We do not focus enough on ultimate beneficiaries i.e. children and families – mostly our evaluations are for programme oriented people and professionals.

How useful were the evaluations for assessing and improving the specific initiatives in question?

Very useful	50%
Generally useful but with some important exceptions	46%
Not often useful	4%
Number of responses:	46

Selected Comments:

General follow-up and implementation of recommendations is sometimes weak.

Sometimes end-cycle evaluation is just done for the sake of it.

Evaluations can be useful if they draw the attention of the concerned ministries towards the commitments made by the Government, and the importance of the evaluation exercise itself as a milestone.

Not all evaluations suggested clearly adjustments based on found shortages. Some evaluations were done at the conclusion of projects so only lessons could be drawn from them for eventual multiplication in the future.

All evaluations led to an expansion of the initiative to more geographic areas.

In case when the evaluations were conducted in the mid-term of projects, they were useful in providing arguments for lobbying and negotiating with Government.

Some evaluations were kept in doors and some planners and managers did not use the results in their planning process.

The evaluation tackled other underlying issues such as quality control, promotion and availability. Some issues that should have been addressed are affordability (price differences between iodized and non-iodized salt) and public awareness on the importance of consuming iodized salt.

Much depends on initial buy into the study. I have found an inverse relationship sometimes in academic quality and ownership and utilisation.

A very negative evaluation may now result in the cancellation of a programme – we are working for it to result in transformation instead.

Some initiatives changed or ended prior to results being available.

Sometimes evaluations do not attract the attention of policy makers, because they have not been involved in the process since the beginning.

In many cases, analyses from evaluations only confirm expected hypotheses, and often do not yield unexpected or new findings.

Useful but not always realistic in regard to availability of resources and knowledge.

Some projects were evaluated on the end of implementation and the end programme cycle. Bad timing – no possibility for change /improvement.

11. How useful were the evaluations for improving UNICEF programming in general in your country, beyond the specific initiative evaluated?

Very useful	20%
Generally useful	67%
Not very useful	13%
Number of responses:	46

Selected Comments:

Mostly there is no analysis in depth and co-ordination among sectors especially for cross cutting issues.

While some evaluations point out problems common to the overall UNICEF programming (e.g. managing NGOs, issues related to the government, etc.), they have not been used or shared as much by others as they could be. It could be due to the fact that these reports are usually not shared outside the concerned project staff.

Depending on evaluation – some recommendations initiated new approaches, e.g. Special Classrooms for Children with Disabilities initiated a pilot inclusive schooling project which was expanded in next years to 20 locations country-wide.

The evaluation was generally useful, as most findings made have been noticed at programme/project implementation level.

The EPI evaluation showed useful lessons that can be learned from the problems of the cold chain, logistics, how to improve on disease surveillance and overall programme management and co-ordination.

Especially the evaluation of HIV/AIDS projects, led to the development of a revised HIV/AIDS strategy paper in which all major stakeholders in the country were involved. This evaluation made also reference to international best practices, which was crucial for the development of this strategy.

Evaluations are always project specific. Little further relevance.

In general in UNICEF, there is a high turnover and a huge workload. So it is really difficult to systematically use such information, unless consultants are hired.

Sometimes the comments mentioned in the evaluation cannot apply in practice.

12. How relevant were the evaluations' lessons and recommendations for UNICEF globally, across several or many countries and regions?

Highly relevant	15%
Generally relevant	75%
Not generally relevant	10%

Selected Comments:

The relevance to other offices could be improved by enhanced mechanisms for sharing – both the preparation of the evaluations and the reports.

However there is a gap between what is theoretically learned and what is practically implemented.

Les leçons tirées et les recommandations portent sur l'amélioration de la mise en œuvre du Programme de Coopération et quelquefois sur les procédures et processus de travail de l'UNICEF.

They could be useful particularly among the countries in the same region having the same type of projects in a similar setting.

Don't know - no feedback from other COs or NYHQs received.

UNICEF/Botswana in collaboration with the Government of Botswana was requested to produce a PMTCT Best Practice report to be shared with other countries.

Mine Awareness and Reintegration are relevant.

This is just assumption because findings and reports were often shared with New York, but the CO cannot determine their reach beyond that level.

Reports were often shared with New York, but the CO cannot determine their reach beyond that.

Major evaluations more useful than very programme-specific ones.

The lessons learnt were relevant for UNICEF globally as they showed that the country has not achieved most of the goals set for the previous decade.

I cannot answer this question due to lack of feedback. However, the HIV/AIDS evaluation was published in the monthly magazine Evaluation and Planning of December 2002 and reactions have so far not been received. The evaluation of Community Based Surveillance was done in collaboration with CDC Atlanta, which may have used it for other countries. The evaluation of Childscope has received attention in the State of the World Children and other countries and donors have visited the projects but whether the recommendations of the evaluation have been utilised is unknown.

We, unlike other countries, have a highly decentralised structure with vibrant society. So sometimes the evaluation lessons and recommendations of this country are not relevant to other countries. At times, findings of some states of this country are not relevant to the other states of this country due to high interstate variations. However, within the country some evaluations have backed up the advocacy work of UNICEF.

13. To what degree did the CO learn from these evaluations how to do better evaluations in future?

A great deal was learned about doing high quality evaluations	26%
There was some learning, but not to the extent that capacity changed significantly	66%
Little was learned about how to do evaluations better	9%
Number of responses:	43

Selected Comments:

Critical lessons learned include the value of wide discussion of TORs; a clear articulation of the purpose of the evaluation; a clear work plan with time built in for incorporation of feedback; and mechanisms for dissemination and monitoring implementation of recommendations.

Les partenaires sont impliqués dès la planification et l'élaboration du Programme de Coopération c'est à dire depuis l'analyse de la situation jusqu'à l'élaboration du MPO. Ils sont impliqués dans l'élaboration du Plan Intégré de suivi et Evaluation pour renforcer leur capacité en planification d'études et évaluations basées sur des critères précis car l'évaluation permet d'améliorer la gestion des programmes, des projets et des activités d'accompagnement afin d'avoir une utilisation optimale des fonds et autres ressources; tirer des enseignements de l'expertise afin d'améliorer la performance, les méthodes et les résultats des programmes de coopération, de fournir des informations permettant d'intensifier le plaidoyer en faveur des politiques, des programmes et des ressources qui améliorent la situation des enfants et des femmes.

Now we have left the CO, much learning left with us.

The CO learned from this evaluation namely how to improve practical modalities of undertaking forthcoming evaluations (need to involve external assistance in the process and evaluations have to be more focused on specific programme components/objectives).

Evaluation was not a priority in capacity building during the examined period.

There is still need for improvement and further capacity building. Good evaluations take time and are generally demanding. For small COs the demand on staff time can be very high.

No significant time and capacity was allocated by this office to manage the evaluations

UNICEF Staff/Counterparts tend to care about the findings of the evaluations but little about the methodology.

There has never been any feedback on our evaluations. In addition, the evaluations that are referred to us as good, or evaluators that are recommended, are not affordable. Most of the officers do not have the skills in developing good projects log frames that make the evaluation a useful exercise. Evaluations often point to known situations that are not easy to change.

We benefited from some feedback received from technical experts outside the country on how we could improve on some of the technical aspects in future evaluations of the same type. We had training on evaluation management, often the weakest part of the process.

14. To what degree did these evaluations improve the capacity of national partners to do evaluations themselves?

A great deal was learned about doing high quality evaluations	2%
Some partners improved the evaluation capacity much more than others.	39%
There was some learning generally, but not to the extent that capacity changed significantly	43%
Little was learned about how to do evaluations better	16%
Number of responses:	44

Selected Comments:

In many cases, there is not enough quality monitoring or guidance from UNICEF during the evaluation is being conducted. Emphasis seems to be put on producing a report, rather than its substance and quality. Furthermore, UNICEF programme officers themselves may not have appropriate capacity to improve the quality of the evaluation.

Le renforcement des capacités des partenaires à porter sur : La planification des évaluations ; Pourquoi faire une évaluation ? Les raisons justifiant une évaluation ; Elaboration des termes de référence ; Formulation des questions spécifiques de l'évaluation ; Comment rédiger un rapport d'évaluation?

Taking the specifics of SIDS into consideration one has to find a good balance to develop national and/or regional specialized capacities for evaluations, looking into cost effectiveness and human resource availability.

Lack of evaluation capacity among national partners is mainly linked to the constraints of low morale among civil servants due to irregular pay and resistance to collaboration among line ministries.

Some partners in the Statistics Department and the Planning Units of Education and Health have improved their capacity to evaluate much more than other partners in the Departments of Social Welfare, Community Development and the Child Protection Alliance, an NGO.

The evaluations were very useful to partners.

This is a huge country and the impact of any single UNICEF supported initiative is relatively limited.

Education counterparts have benefited greatly from the two Interactive Learning Evaluations.

In the highly unstable context of this country, there is no unified government or national level counterpart and UNICEF works with de facto authorities and mostly local communities. Therefore, this question is not really applicable.

The quick turn over of the staff is one of the major constraints. Always having new comers. There are many research institutes who can perform evaluations. The question is quality. Counterparts were involved but due to other priorities, it did little impact to capacity development.

There is a blame culture here and evaluation is seen generally as inspection and criticism = this will take years to change.

15. To what degree did these evaluations improve the programming or policy capacity of national partners?

A great deal	24%
To some degree, but not significantly changed	64%
Little	11%
Number of responses:	45

Selected Comments:

National partners still need capacity building in programming and policy development.

Difficult to answer this question in general as every evaluation is specific. For the specific intervention of Community Based Surveillance the capacity for national partners has certainly improved.

Several of our partners now pay greater attention to evaluation and evaluation findings.

In our national context it is difficult to have an evaluation of importance, which can significantly influence the programming, or policy capacity of national partners. However, at the state level, there have been few instances where the evaluations combined with advocacy and available local opportunities triggered change.

Planning is usually done on a higher level.

Les résultats des évaluations ont permis de tirer des leçons d'être plus réaliste dans la formulation des objectifs, de formuler des objectifs basés sur des résultats, de mieux se focaliser dans une zone bien définie pour mesurer l'impact du programme sur les bénéficiaires et de mieux définir et retenir des indicateurs clés permettant de suivre et de mesurer les progrès accomplis.

The evaluations fed into the pool of general knowledge, but there was no obvious influence on programme policy and development.

In many cases the evaluations provided the basis for future planning – but as planning itself is rather weak, the national capacity will take some time to build up.

16. In your opinion, did these evaluations, together, give an accurate picture of the quality of UNICEF’s interventions?

Yes	19%
Yes, with significant qualifications	81%
A comprehensive picture?	
Yes	0%
No	100%

Selected Comments:

Results from evaluations do not always show the qualitative achievements of our interventions. For example, children may be enjoying the schools more after our project intervention, but this fact does not necessarily be reflected in the quantitative outputs such as enrolment or academic achievement.

The lack of a complete picture is mainly due to the lack of updated data in key social sectors.

Evaluations are done on specific projects and do not represent the entire programme.

The quality standards are not observed for most locally recruited consultants.

Nos partenaires du gouvernement, les agences du système des Nations Unies, les partenaires bilatérales, multilatérales, les ONG et la société civiles apprécient beaucoup les rapports d’étude et d’évaluations de l’UNICEF car riche en données bien désagrégées par découpage administratif, par sexe sur les enfants et les femmes pour une prise de décision et planification des actions.

Evaluations were very limited in scope and the quality was mixed.

It is important to note that UNICEF program interventions in this country are very different compared to other country programs. This is a country in transition where the government is fully responsible for the program. UNICEF role is limited. For example, we could not claim that salt iodization is a UNICEF intervention. It is a government intervention. UNICEF sold the idea to the government.

We only evaluated some of the interventions so not comprehensive.

17. Was the independence of the evaluators adequately protected?

Almost always	83%
Generally, but with some important exceptions	15%
Frequently not	2%
Number of responses:	46

Selected Comments:

The national consultant hired for this evaluation did not face major protection constraints.

Independence was generally OK except maybe for a few occasions with some political difficulties.

Not sure.

The evaluators are not allowed to be independent in terms of sensitive issues and travelling in this country.

Oui, puisque les évaluations et études se font sous la supervision du gouvernement et valider par ce dernier par le gouvernement lors de l'atelier de restitution.

The evaluators are the staff of the ministries responsible for salt iodization, hence theoretically there might be some conflict of interest arising from the fact that the same people salt promote salt iodization were trying to assess how widely households are consuming iodized salt.

18. In 2002, does the country office record its decisions on each recommendation of an evaluation?

No	74%
Yes.	26%
Number of responses:	46

Is this a change of practice since 200-2001?

Yes	23%
No	77%
Number of responses:	22

If the CO does record a decision on each recommendation, how does it do so?

The recommendations were incorporated into the annual project plans of actions for implementation.

We do not systematically record the decisions or follow-up taken from recommendations of an evaluation as such, through a checklist but do certainly follow-up on several recommendations, especially those that seem realistic and feasible to act on. The way to record decisions taken are available in review documents of projects, and in minutes from feedback sessions when evaluation findings were disseminated.

Done in the processes of 2002 Mid-Term Year Review –i.e., study validation/project review/programme revision meetings.

Yes, the recommendations were systematically reviewed as part of the annual review process in order to be implemented in the coming year, and in preparation for the new county programme. Moreover, there was a special evaluation dissemination meeting convened with partners, working group sessions were held on the various categories of recommendations and proceedings were developed, recording decisions on the recommendations of the evaluation.

Each evaluation report has a chapter with the recommendations both for UNICEF and national partners regarding improvement of future programme activities.

The monitoring and evaluation process (IMEP) has been systematic in the programmes that integrate UNICEF's cooperation in the country, actively incorporating the operational counterparts in the evaluation process.

Country Programme and annual PPAs.

Usually recommendations are incorporated in the next PPO or PPA.

Using findings and recommendations in the planning process.

The recommendations were incorporated into the annual project plans of actions for implementation.

The CP programme area PPA.

Individual changes are discussed at Programme Meetings, which are held every two weeks. Follow up is also done with Counterparts and PER evaluations are also used.

Any follow-up is left to the Officer conducting the Evaluation.

The Office Programme Plans of Action, and Annual Management Plan, have incorporated the actions to be taken in follow-up to the evaluation. Moreover, those recommendations to be taken into account in the preparation of the new CPP have been extrapolated to be reviewed in the course of the planning process.

Evaluations and findings from evaluations are taken into consideration when planning a new period.

Main decisions are reflected in annual Project Plans of Action.

In accordance to each recommendation generated by the carried out evaluation, it is incorporated within the actions of the evaluated programme and it is established for its implementation.

Follow-up actions were included in the PPAs and discussions were held with counterparts for further improvement.

Relevant findings of the evaluation do greatly contribute to the designing of the annual plans of action.

Please indicate if this is a change in practice since 2000/2001.

Yes 14%

No 76%

Number of responses: 21

Is an action plan prepared after each evaluation?

Yes	40%
No.	60%
Number of responses:	42

If an action plan is not prepared, are there other mechanisms for follow-up?

No	50%
Yes.	50%
Number of responses:	16

If there are other follow-up mechanisms, please explain what these are.

Plans of action were developed after some evaluations including ---. This is becoming more frequent.

By compiling recommendations and its follow-up status in a matrix, and reviewing them regularly.

These Evaluations were used to guide our future support to projects to ensure effectiveness and relevance. While there is no action plan prepared by the Office, partners are required to reflect recommendations in project implementation, of which continued funding is premised.

In several instances, even when action plans were not prepared, recommendations were acted upon, such as having Child Friendly Hospitals and not just Psychosocial Care for Hospitalised Children.

Although action plans are not prepared after each evaluation, action plans were prepared after the evaluation of the PMTCT programme on how to implement the scaling up of the programme with consideration of the findings of the evaluations. Action plans to develop the Infant and young feeding policy, and the accelerated plans to promote and protect the breast-feeding were developed after the evaluation of the infant and young child feeding practices.

Discussion with counterparts on results and follow up actions.

19. In your opinion, were the evaluations worth the resources invested?

Yes	93%
No	7%
Number of responses:	43

Selected Comments:

They were worth the investment, but maximum value was not always achieved. Sometimes the quality was not adequate to base an action plan on the report. Sometimes solid evaluations with recommendations are not sufficiently followed up.

However, mechanism for follow up in practice must be improved.

If not all, at least some of the evaluations were worth the resources invested. But more focus needs to be given on the rationale for conducting these evaluations and give more attention to the selected few major evaluations, method, choice of consulting agency, quality monitoring, and sharing the results.

Having in mind everything written above, the evaluations were worth the resources. We might insist in the future on a wiser use of the results and agreement of all stakeholders with regards to implementation of an action plan.

The evaluations were worth the resources, but the results should have been used more wisely.

Important lessons in the process even if the outcomes not altogether useful.

More were worth the resources invested than were not. The bigger the evaluation, the more worthwhile in general.

Evaluations are a way to document and share experience within the Organization, and to plan better future activities.

In many cases we would not have known how to proceed without the Evaluation.

No significant resources have been invested in good evaluated work. We need much more investment.

Most of the times But not all of the times.

As learning process we can say yes, but with stereotype recommendations we can say no.

Objective evaluation on the project provided critical assessment of success and failures. As it is difficult to have travels to programme sites due to travel restrictions by air, intensive evaluation helps a lot to understand situation and end-results.

To some extent yes, because some of the recommendations are not much relevant with real situation. Some evaluations are not so cost effective.

Not always. Some evaluations ended up destroying office teams. Utilisation was low when they became too academic with insistence on high quality product instead of process.

In general, yes but we have no standard to compare too. In general UNICEF can get this done more cheaply than say the WB so definitely better value than other international agencies.

20. What one or two things would, if implemented, most improve evaluation practice at the UNICEF CO level?

Institutionalize the wide sharing of TORs, work plans and draft reports, to include more national, regional and global partners. This would require a mechanism to ensure a speedy turnaround. Second, UNICEF needs a mechanism for ensuring follow-up implementation of recommendations.

At UNICEF CO level, evaluation is to be conducted by a team, composed not only of PO in charge of program, but in collaboration of experts from planning and policy sector. Government institutions (policy makers) must be part of the process since the beginning.

Establishing a quality monitoring mechanism throughout the evaluation (i.e. from deciding whether to conduct an evaluation or not, developing TOR, methodology, field survey, in-depth analysis, report writing, etc.). The information regarding evaluations and studies should be centralized and put on the office Intranet/LAN to make it accessible. It is also necessary to keep record and catalogue of various important documents for use by many. Networking among partners in and outside the office (including non-UNICEF partners) to share findings and lessons learned from their experiences would be beneficial.

Structured Monitoring and Evaluation Programme that will constantly review objectives vs. outcomes and make adjustments accordingly. Adequate staffing to provide the appropriate level of technical assistance.

Better technical knowledge on methodologies and techniques of evaluation. More quality and cost effective consciousness.

Regular distribution of the good evaluation reports done in UNICEF worldwide. Updated best practices while managing the evaluation. Updated roster with internal / UNICEF and external experts in evaluation in certain fields. Outline of Terms of Reference / format with tips what is important.

Greater involvement of partners in evaluation.

Improve the utilisation of evaluation results. Improve sharing of evaluation findings across programmes.

HQ funding for evaluations; Use of independent, qualified and competent evaluators; Detailed TORs; Training for CO staff in conducting/managing evaluation processes.

Participatory process of identifying relevant evaluations/themes for the year rather than relying on sectoral proposals; Identifying follow-up/monitoring mechanisms or plans of actions for implementation of evaluation recommendations.

Prepare and monitor an action plan after each evaluation.

Participatory process of identifying relevant evaluations/themes for the year rather than relying on sectoral proposals. Identifying follow-up/monitoring mechanisms or plans of actions for implementation of evaluation recommendations.

Training staff on conducting evaluations. Hiring competent and well-reputed consultants to do the evaluations. Prepare a coherent and relevant IMEP. Give evaluations of key interventions top priority. UNICEF HQ database should have full e-copies of best examples of good practice evaluations.

The establishment of an evaluation plan based on relevant indicators for the various programme components and covering the five year period, jointly with major Government counterparts.

Capacity building in evaluation and the setting up of National Association of Evaluators to improve and nurture the culture of evaluation particularly among partners.

1) Increase publication of good evaluations and clarity on our publication policy. Often we do not publish evaluations officially, but disseminate rather as a simple photocopied document, internally and within the country. Though this reduces costs it is not very attractive to read for other parties than those directly implicated. We were quite happy when the regional office in ESAR proposed the HIV/AIDS evaluation for publication in the monthly magazine "Evaluation and Planning". This gives importance to good evaluations and disseminates results to a much wider audience. It also gives an incentive to consultants to do a great job if these possibilities exist.

2) To increase capacity of evaluations, each regional office could request what are the main themes for evaluations and support these evaluations in different countries by helping to identify or proposing excellent evaluators and setting up excellent TORs. It would also be good to ensure that M&E officers are now and then given the opportunity to participate in an evaluation. This practice would allow UNICEF Country offices to increase capacity. At the same time, thematic evaluations in several countries in a region (or even globally) would allow UNICEF to make some good publications on relevant topics that can be widely disseminated.

3) It is difficult to find the right person, with adequate knowledge of the theme and analytical skills for conducting evaluations. The Country Office should establish a list of possible evaluators, locally available. For international consultants, a list of consultants at regional and Headquarter level would be convenient. It is important to have contact persons for each consultant proposed, which have had direct working experience with them. The list of evaluators listed under the African Evaluation network doesn't have this type of direct experience contact and it is therefore difficult to judge what is the actual track record of consultants listed here.

4) Timeliness of submitting evaluation reports is a major problem. Despite a clear and realistic timeline in the TOR, consultants often have other commitments and time becomes difficult to control, even with penalties in payment after delays.

IMEP and a competent M&E Officer.

Someone who could spend more time on M&E of Evaluations.

Circulation of best practices in evaluations with examples from each region. Commitment to prepare IMEP and adhere to IMEP. Hire M& E Officer in CO. At the moment we have a focal point but this is clearly insufficient.

1. Affiliation with local research institutes that have evaluation capacity. Partnerships of some of these institutes with international evaluation research agencies are required.

2. Resourcing the evaluation function at Country Offices and at Hqrs.

3. Capacity building training to better manage the evaluation function and examples of good evaluation work.

Our comments apply to UNICEF country office practices in general.

Evaluation protocol & design – especially those for summative evaluations – must be drawn up at the start of the programme. Usually programme officers tend to decide on such an evaluation only towards the end of the programme (number 8 in the figure below, and more towards its “weak” side).

The evaluation capacity of UNICEF staff is one area that requires urgent improvement within the organisation. Generally, recruitment of staff against country level M&E posts is not based on technical capacity and/or experience in evaluation.

In many programmes at country level, UNICEF’s principal inputs are technical and policy contributions: if sound, such contributions are usually based on research and evaluation. The role of country level UNICEF M&E staff is therefore crucial, and such staff should provide “added value”— to other programme officers, and to other development partners.

If the evaluation function is to become a crucial part of the programme, the organisation needs to institute some accountability mechanisms, closely linked to the quality of programming. Currently bad programming or good programming, or managing the evaluation function well or poorly does not seem to matter.

All Evaluations TORs of Evaluations and of their consultant should be shared with the M&E officer (or the Programme Co-ordinator if the CO does not have an M&E officer). The programme section involved and the M&E officer should evaluate all Evaluation studies. Implementation of recommendations, if they are relevant, appropriate, functional and effective/useful, should be followed up closely by the programme section and the M&E officer (or the Programme Co-ordinator if the CO does not have an M&E officer).

A binding mechanism must to be developed, at least at the CO level, to ensure optimizing returns from evaluations, i.e. maximum returns/benefits from the money spent on the evaluation and the implementation of its recommendations.

Have a clear plan for the follow up to the evaluation; this is not always the case. Reduce petty bureaucratic responsibilities faced by project officers, in order to give them sufficient time to REALLY read and reflect on various drafts and discussion. All too often things are quickly cooked because there are a million things we are involved in.

Have the staff learn the differences between programme/project evaluations and the situation analyses or baseline studies. Check closely and document the implementation of “recommendations” that are made in the particular evaluation studies. Establish this habit as a part of all Evaluation practices.

Problems: Some COs are heavily dependent on emergency funds, which are erratic, can never be planned and are always short-term. In such conditions, it is not always possible to plan programmes as full circles and to estimate the duration. In addition, the evaluations are sometimes conducted according to requirements of the donors, and very little initiative is left to the CO.

Suggestions:

It would help if standards for Terms of Reference were developed and a checklist provided so that all responsible staff have to undergo the same process;

It would also be beneficial to conduct more mid-term formative evaluations, rather than end-of-project summative ones, that have the purpose of final reports.

Training and orientation (by regional offices and Headquarters) on development of terms of references for evaluations, and in providing inputs to the development of evaluation methodologies and protocol.

Continued and enhanced involvement and participation of Regional Offices (from a technical oversight perspective) and Headquarters (from a policy guidance perspective) in supporting Country Office evaluations would be highly beneficial.

Evaluation be checked with the quality standards (utility, feasibility, accuracy and property Plan of actions to implement recommendations.

Adequate and wide sharing of information starting from the point of planning for Evaluation up to the stage where the evaluation is conducted dissemination of the results.

Establishment of mechanism of quality control for the entire process of conducting an evaluation (this should start from the preparation of TOR, selection of consultants and follow-up of recommendations)

Clearly defined terms of reference and monitoring system for follow-up of recommendations.

Training for all staff of the Office in monitoring and evaluation.

Have a clear definition of a more participative methodology to in the evaluations all the actors involved in the projects.

Planning in advance a schedule for evaluations, so as to make all people involved about the need for their participation.

Need training on M&E, especially research skills. Need an intra-regional networking to share experiences and mechanism to record lessons learnt systematically.

More involvement of different programme sections, partners and target groups in the evaluation process.

More structured mechanisms to analyse, “digest” and integrate evaluation findings and recommendations in the programme management and implementation process.

Training: Support from Regional Office.

In order to reinforce the evaluation culture at CO level, it is recommended to carry out periodic training workshops that up-dates the personnel on evaluation methodologies and that stimulates the monitoring and evaluation ability jointly.

Joint evaluation (external and internal) could be useful for learning process and improving programme implementation.

Training to enhance capacity of counterparts and UNICEF staff for more analytical evaluation.

Officers from other CO's with similar programmes could be included in the team.

Learning from evaluations of one section could be shared with another.

Evaluation should be more widely disseminated.

In our country, during an evaluation on training supported by UNICEF, design of the evaluation, work process of data and information gathering, analysis procedure and periodically progress status of the evaluation are very good examples to improve the capacity of the evaluation at the CO. However, national programme officers could not participate efficiently (apart from facilitation in logistic issues for the consultants) in the process of the study because the national officers were not able to find enough time to participate in the study.

Develop matrix on evaluation/finding which action plans for follow-up and monitoring through programme meetings and other review forums; recording and implementing relevant decisions;

More partners/stakeholders should be involved.

Raise more awareness of the importance of evaluation practice in programme management cycle.

Development of plan of action of the recommendations from evaluations.

A thorough review of the evaluation's (or study's) framework, assumptions, and tools by the concerned UNICEF technical unit at the HQ. Not all Regional Offices have the capacity to provide such technical advice for all areas of evaluations conducted at country level.

Direct interest by senior management and review of IMEP by CMT as a part of regular agenda.

High Professionalism of the Consultants; Rigorous TOR.

Some guidelines on how much to invest.

Une bonne planification des études et évaluation ; Terme de référence bien élaboré avec des objectifs clairs et précis, les utilisateurs des résultats, les méthodes de collecte ; Etre plus analytique que descriptive dans la rédaction du rapport, bien désagrégées les informations ; Faire participer les bénéficiaires au processus de planification et restitution des résultats et non seulement à l'étape administration des questionnaires.

The office requested a full time evaluation post, which was not, approved. We consider this essential. Additional training is needed.

APPENDIX 3: MODEL TABLE OF CONTENTS FOR AN EVALUATION OF STRATEGY

The following is a model Table of Contents for a splendid evaluation report. Other report structures are possible, of course, and in a particular case might be preferable. However, unless there is a good reason to vary, standardisation makes the report more accessible to managers, and ensures that important topics are not inadvertently omitted.

Table of Contents

Title page [title, author's name and address, date, report number]
Disclaimer
Letter of Transmittal
Authorities and sign-offs
Evaluation Executive Summary
Table of Contents [with list of appendices]

1. INTRODUCTION

Purpose of this Evaluation
Scope and Limitations
Ethics and Independence
Methodology
Analysis Plan
Data Collection and Sampling Plan
Summary of Previous Evaluations and Pertinent Research
UNICEF Research
Other Pertinent Research
Summary of Start-up Workshop
Stakeholder Perspectives
Evaluation Issues

2. RATIONALE – A RIGHTS-BASED PERSPECTIVE

The Rights Situation [Cultural, social, political, and economic context]
Purpose, Components and History of the UNICEF Intervention
Role and Importance of the Intervention in the Country Program
Host Country Stakeholder Capabilities and Priorities
Logical Framework
Related Activities by other Agencies

3. EXPENDITURE ANALYSIS

UNICEF Expenditures
Other Party Contributions in Cash or in Kind
Audit Results or Requirements

4. ACTIVITIES AND OUTPUTS

Organisation, Roles, and Responsibilities
Implementation Schedule and Achievement
Planned, Actual, and Projected Outputs
Management and Financial Control
Capacity, Training, and Development
Co-ordination with Partners
Replicability

5. BENEFITS ANALYSIS

Improvement in the Rights Situation
Welfare of Children and Women [economic, social, political]
Gender Effects
Environmental Effects
Overall Value for Money

6. SUSTAINABILITY

Is Sustainability Necessary?
Cost Recovery/Revenue Generation?
UNICEF Commitments Required
Other Sources of Funding
Sustainability Prognosis

7. ALTERNATIVES

Alternative Types of Interventions? [Direct service, volunteer/cooperative, or market-based]
Scope, Focus or Timing Alternatives?
Partnership Alternatives?
Management Alternatives?
Funding Alternatives?

8. LESSONS LEARNED

Development Lessons
Lessons for the Next Phase
Other Interventions in this Country
Similar Interventions in other Countries
UNICEF Operational Lessons
Lessons about Conducting Evaluations
Summary Observations of the Exit Workshop

9. RECOMMENDATIONS

UNICEF-actionable Recommendations
Implementing-Agency-Actionable Recommendations
Partner-Actionable Recommendations
Funding Recommendations

10. ACTION PLANS

Country Office Action Plan
UNICEF HQ Action Plan
Partner Action Plans

APPENDICES

Data Tables
Figures
Terms of Reference
Mission Itinerary, Persons Interviewed, and Field Observations
Project Photographs
Data Collection Instruments
Bibliography
Endnotes

APPENDIX 4: MODEL TABLE OF CONTENTS FOR A PERFORMANCE EVALUATION

The following is a model Table of Contents for a splendid evaluation report. Other report structures are possible, of course, and in a particular case might be preferable. However, unless there is a good reason to vary, standardisation makes the report more accessible to managers, and ensures that important topics are not inadvertently omitted.

Table of Contents

Title page [title, author's name and address, date, report number]
Disclaimer
Letter of Transmittal
Authorities and Sign-offs
Evaluation Executive Summary
Table of Contents [with list of appendices]

1. INTRODUCTION

Purpose of this Evaluation
Scope and Limitations
Ethics and Independence
Methodology
Analysis Plan
Data Collection and Sampling Plan
Summary of Start-up Workshop
Stakeholder Perspectives
Evaluation Issues

2. OBJECTIVES AND TARGETS

The Performance Framework
Targets and Commitments

3. EXPENDITURE ANALYSIS

UNICEF Expenditures
Other Party Contributions in Cash or in Kind
Audit Results or Requirements

4. ACTIVITIES AND OUTPUTS

Accountabilities [Organisation, Roles, and Responsibilities]
Implementation Schedule and Achievement
Planned, Actual and Projected Outputs
Management and Financial Control
Capacity, Training, and Development
Co-ordination with Partners

5. BENEFITS ANALYSIS

Objectives Achievement [Improvement in the Rights Situation]
Welfare of Children and Women [economic, social, political]
Gender Effects
Environmental Effects
Overall Value for Money

6. SUSTAINABILITY

Is Sustainability Necessary?
Cost Recovery/Revenue Generation?
UNICEF Commitments Required
Other Sources of Funding
Sustainability Prognosis

7. LESSONS LEARNED

Development Lessons
Lessons for the Next Phase
Other Interventions in this Country
Similar Interventions in other Countries
UNICEF Operational Lessons
Lessons about Conducting Performance Evaluations
Summary Observations of the Exit Workshop

8. RECOMMENDATIONS

UNICEF-actionable Recommendations
Implementing Agency-Actionable Recommendations
Partner-Actionable Recommendations
Funding Recommendations

9. ACTION PLANS

Country Office Action Plan
UNICEF HQ Action Plan
Partner Action Plans

APPENDICES

Data Tables and Figures
Terms of Reference
Mission Itinerary, Persons Interviewed, and Field Observations
Project Photographs
Data Collection Instruments
Bibliography
Endnotes

APPENDIX 5: TERMS OF REFERENCE

The Evaluation Office (EO), among other offices, is responsible for monitoring the quality of UNICEF-sponsored evaluations. It is also responsible for maintaining an organisational database of evaluation and research and seeking to strengthen the UNICEF-supported evaluation as part of a comprehensive performance management system (E/ICEF/2002/10).

UNICEF is a very decentralised organisation, including in its evaluation work, the bulk of which is done at the Country Office (CO) level. Around 1,000 evaluations and studies are typically reported each year. Regional Offices and the EO itself add only a small number of evaluations and studies to these total figures.

In 1994, the then Evaluation and Research Office carried out a systematic cross-sectoral review of UNICEF-supported evaluations and studies.¹⁰⁰ The Evaluation Office¹⁰¹ and Regional Offices have since carried out thematic Desk Reviews, some of which have included a meta-analysis of the quality of evaluations. However, it has been close to a decade since an overall assessment has been made.

At the same time, in the last few years, the EO has adjusted its strategy to improve the quality of evaluations, distinguishing more clearly approaches in relation to evaluation at policy, country programme and programme component levels. A number of basic but important initiatives to strengthen evaluation at the latter two levels as managed by COs have just or are about to come to fruition. The EO has worked to improve clarity on what is expected of COs with regard to country programme and programme component-level evaluation through revisions to policy and procedures (1999 and 2002). The EO also continues to work on the development of a key tool to help COs in the management of these responsibilities — draft guidance on the Integrated Monitoring, Evaluation and Research Plan have been widely circulated and training has been undertaken in four regions.

The Office has also developed training materials to help staff in building the necessary knowledge and skills in managing programme monitoring and evaluation activities — a draft CD-ROM was distributed in 2001, will be finalised and reissued this fall and followed by a Training of Trainers in 2003. Other work is ongoing to further clarify country programme level evaluation, its relation to the CO-managed Mid-Term Review and evaluations and reviews at the level of the wider United Nations Development Assistance Framework.

The EO is at a stage where it can soon expect to see the effects of some of the above efforts on the quality of evaluation at CO level. At the same time, as some of the above activities wind up or are handed over to others, the EO must now begin to refine and develop further strategies to strengthen evaluation. EO must do so taking into account the diverse strategies pursued by the different Regional Offices and the important contributions of other headquarters units. As a key input to this process, a commonly accepted and well-founded assessment of the quality of UNICEF-supported evaluation is needed.

¹⁰⁰ Victora, C. et al. (1994) A Systematic Review of UNICEF-supported Evaluations and Studies, 1992-93.

¹⁰¹ And its previous incarnations, the Evaluation Section in the Division of Evaluation Policy and Planning (1996-2000); the Evaluation and Research Office (prior to 1996).

Purpose

The proposed meta-evaluation is intended to provide an overall assessment of the quality of UNICEF-supported evaluation and research at CO level — currently and as it has evolved since 1994.

The key and immediate users of the meta-evaluation will be the Evaluation Office and the Regional Offices. The meta-evaluation will feed into a reassessment of evaluation strategies to strengthen UNICEF-supported evaluation, focusing on CO capacities. The Evaluation Office will in turn likely use the results of the meta-evaluation in communicating with the Global and Regional Management Teams to establish a common understanding of the situation and galvanise different parts of the organisation around complementary strategies and actions.

The meta-evaluation will also provide a baseline against which the effects of recently introduced policies and guidance materials, tools, and training strategies can be measured in a few years' time.

Scope and Focus

The meta-evaluation will cover UNICEF-supported evaluations at the CO level, where the bulk of evaluation work is undertaken. The quality of evaluation in 2000 and 2001 will be compared as much as possible with that of 1992 and 1993, as was covered in the 1994 meta-evaluation.

In designing this meta-evaluation, the analytical questions to be addressed and the corresponding methodology have to accommodate this comparison with the findings of the previous 1994 meta-evaluation. At the same time, new questions are posed by the evolution of UNICEF guidance on evaluation in the years since and current concerns regarding the evaluation function.

The following were the initial set of analytical questions:

- How does the quality of evaluation reports compare to current evaluation standards¹⁰² (in particular those aspects that can be measured from the evaluation report, such as most utility standards, selected propriety standards and most accuracy standards)?
- How adequate is the presentation of evaluation and study reports, specifically the executive summary, introductory/background materials, methodology, findings/conclusions, recommendations, and lessons learned?
- How well do reports distinguish implementation (inputs, process milestones, activities completed), from outputs and outcome/impact and to what degree and how well are each addressed?
- What is the overall impression of the quality of evaluations?
- How does the quality of evaluations and studies compare?
- At what level of analysis are evaluations focused — country programme, programme components, projects or activities? How does this affect their utility and the quality of their interpretation of different levels of results?

¹⁰² See Joint Committee on Standards for Educational Evaluation. (1994) The Program Evaluation Standards and similar draft African Evaluation Guidelines 2000 under development by the African Evaluation Association.

- To what degree have evaluations and studies contributed to capacity building of national partners and/or UNICEF COs?
- To what degree and in what roles are key stakeholders involved in evaluations and studies, especially primary stakeholders?
- To what degree and how has the overall design and management of evaluations addressed the issue of impartiality?
- To what degree have evaluations and studies resulted in follow-up actions at country level?
- How does the overall assessment of evaluations compare with previous assessments, including the 1994 meta-evaluation and other thematic meta-analyses? (See "Information sources" below.)
- What do evaluations and studies examined suggest about the accuracy of overall figures on reports in Evaluation Database tracking systems¹⁰³?

Information Sources

In addressing the questions for analysis outlined above, the meta-evaluation will draw from the following existing sources:

- Actual evaluations and studies received by the Evaluation Office,
- Data from the UNICEF Evaluation and Research Database (managed by the Evaluation Office) as well as a complementary EO system of tracking evaluations and studies reported by COs in annual reports,
- A small number of internal studies reflecting on the quality of UNICEF-supported programme evaluation, including:
 - Victora, C. et al. (1994) A Systematic Review of UNICEF-supported Evaluations and Studies, 1992-93.
 - Chapman, David. (2001) A Review of Evaluations of UNICEF Education Activities (1994-2000).
 - Don Paul and Stein-Erick Kruse. (2000) UNICEF Evaluation Database: A Learning Tool.
- Key policy and guidance references in relation to evaluation, including:
 - The Medium Term Strategic Plan 2002-2005.
 - The Executive Board document "Report on the Evaluation Function in the Context of the Medium-Term Strategic Plan" (UNICEF Executive Board document E/ICEF/2002/10, 11 April 2002).
 - Sections of the UNICEF Programme Policy and Procedures Manual 2002.

Meta-evaluation Methods/Process

The meta-evaluation will draw primarily on new data from the systematic assessment of a sample of evaluations and studies using a standardised tool. This may be complemented by an

¹⁰³ The Evaluation Database itself, as recently revised, does not include a record of all evaluations and studies. COs select what they submit to the EO for entry into the Evaluation Database. However, COs are required to send all evaluations to the EO. In managing the Database, EO does maintain a simple but critical system for tracking the total listing of evaluations and studies reported in the Annual Reports and comparing this to reports submitted to the EO and the Evaluation Database.

e-mail survey of COs, especially on the issue of evaluation follow-ups. Standard quantitative and qualitative data analysis will be used for the results of the assessment of evaluations and studies as well as for survey results. The final analysis will also draw on documentary review of previous meta-analyses and organisational guidance, as well as consultation with Evaluation Office and selected other colleagues as a reference group.

Overall Process

The meta-evaluation will include three phases. A preparatory phase will involve initially only the Team Leader and Evaluation Office staff and will include:

- A thorough review of previous meta-analyses and organisational guidance to provide context for the above questions for analysis;
- The review and further development of questions for analysis as necessary;
- The final design of a sampling scheme for evaluations;
- The development of tools for the assessment of individual evaluations and corresponding database design;
- The design of an e-mail survey (sampling and tool) if determined advantageous; and
- Final selection and preparation of the full meta-evaluation team.

The actual assessment phase will engage the full meta-evaluation team in applying the tool developed to the sample of evaluations. During this phase, any e-mail survey of COs will also be applied and processed.

The analysis and reporting phase will include:

- Quantitative and qualitative analysis of data;
- Preparation of a draft meta-evaluation report;
- Circulation of the draft report to Evaluation Office and selected other colleagues for comment; and
- Preparation of the final report.

Sampling of Evaluations and Studies to be Assessed

The sample of evaluations and studies to be assessed will be drawn from reports produced in 2000 and 2001 and submitted to the Evaluation Office. Considerations guiding sampling choices include the following:

- Evaluations need to be distinguished from studies.
- The total number of reports actually submitted to the EO is still low.
- It is expected that among the evaluation and study reports submitted to the EO but not included in the Evaluation Database, a significant number may be misclassified as to whether they are evaluation or studies or as to whether they fit in either category. Screening will be necessary after sampling to weed out reports that are incorrectly classified. (The 1994 meta-evaluation found 20 percent of reports to be classified incorrectly.)
- The choice of sample size must balance issues of credibility with cost.

Limitations

It is acknowledged that the conclusions of the meta-evaluation will be limited in that only those evaluations available in NYHQ will be sampled. It is safe to assume that reports of extremely bad quality are more likely to have been excluded from the sample. Assessment tools will also have their limitations; in particular it is recognised that some characteristics of good evaluation cannot be assessed based on analysis of the evaluation report alone. A poor quality evaluation report does not necessarily mean that the evaluation process and results were of poor quality.

Rather it reflects the ability of the evaluators to write and present the process and results well, and the relative importance the UNICEF evaluation managers give to getting the report done well. The Evaluation Team will identify further limitations.

Evaluation Team Composition

The meta-evaluation team will consist of Dr. Kenneth Watson as Team Leader and four team members: Anne Perkins, Julia Paton, Andre Bernier, and Zahra Boodhwani.

Accountabilities

The meta-evaluation will be carried out by a team, under the guidance of the Evaluation Office, with one Evaluation Officer, Kate Alley, as a designated focal point. Team Leader will be responsible for the overall project, including:

- Further development of meta-evaluation design;
- Development and testing of necessary tools, including for the assessment of evaluations and studies as well as any e-mail survey of COs;
- Development of a database for the assessment of evaluations and studies (see specifications on products below);
- Final selection, orientation, training and management of other meta-evaluation team members in their contributions;
- Systematic and rigorous assessment of evaluations and studies;
- Implementation of any e-mail surveys;
- Data processing;
- Quantitative and qualitative analysis;
- Preparation of draft and final meta-evaluation reports.

Team members will be responsible for:

- Systematic and rigorous assessment of evaluations and studies applying tools developed by the Team Leader;
- Data-entry;
- Contribution to qualitative and quantitative data analysis.

The Evaluation Office will be responsible for:

- Review and approval of meta-evaluation design and data collection tools;
- Assistance in identifying potential candidates as members of the evaluation team, specifically assistance in identifying individuals with the necessary knowledge of UNICEF;
- Final approval of the selection of evaluation team members;
- Assistance in accessing EO records tracking evaluations and studies submitted, for use in sampling;
- Assistance in accessing and reproduction of EO evaluation and study reports;
- Distribution of any e-mail survey to UNICEF COs and tracking and forwarding of responses received;
- Facilitating consultation with a reference group (see below) and consolidation of their comments;
- Approval of the final report.

A Reference Group will be formed comprising the EO professional staff, as well as selected Regional M&E Officers and headquarters colleagues. The Reference Group members will be responsible for providing input on the analytical questions to be addressed by the meta-evaluation and the draft report itself.

Products

The final product is a meta-evaluation report presenting findings and conclusions on the quality of evaluations and studies carried out at CO level, addressing the analytical questions as outlined in the final TORs, proposing recommendations on future efforts to strengthen UNICEF-supported evaluation practice and the functioning of the UNICEF evaluation system, and where possible presenting lessons learned. The executive summary must follow the specifications for entry in the UNICEF Global Evaluation and Research Database in the attached Technical Note.

The final report should be provided in hard-copy (1 copy) and electronic version in Microsoft Word 97. A style guide will be provided. Raw data from the assessment of evaluations and studies will be provided in electronic format in a database. Choice of database software must be taken in consultation with the EO to ensure compatibility. All electronic files will be submitted on a CD.

Working Arrangements

Team members will be expected to provide their own office space and working equipment. At certain stages of the work it will be expedient for the team leader and later team members to spend some time in UNICEF offices, for which space will be provided. Any daily subsistence allowance and travel expenditures must be accommodated in the total lump sum available for the project (see below).

Reporting/Supervision

The team will work under the overall supervision of Kate Alley, Project Officer, Evaluation Office.

APPENDIX 6: UNICEF RESPONSE TO SPECIFIC RECOMMENDATIONS

The following text was prepared by UNICEF's Evaluation Office with the input of key stakeholders in Regional Offices (ROs) and certain Country Offices (COs). The meta-evaluation contains almost 50 specific recommendations, which are grouped and prioritised here by UNICEF.

OVERALL STRATEGY

Recommendation [7.1.1] UNICEF needs to improve the quality of the bottom third of evaluations to ameliorate evaluation risk.

Recommendation [7.2.2] UNICEF should aim to achieve excellence in some evaluations of intervention strategy each year.

Recommendation [7.2.3] UNICEF should formulate a medium-term "evaluation quality improvement plan" as the action plan arising from this study.

Partially Accepted. A Plan of Action to Strengthen Evaluation in UNICEF 2004-2006 has been developed with a focus on improving the bottom one half of evaluations and achieving excellence in evaluation for selected priority themes.

Recommendation [7.1.2] UNICEF should do fewer evaluations, but should achieve adequate quality in all it does do.

Partially accepted. UNICEF considers that the focus should stay on improving quality rather than reducing numbers of evaluation. UNICEF is and will continue to promote better CO planning, using the Integrated Monitoring and Evaluation Plan, and has noted that CO planning of evaluation does tend to be over-ambitious.

Recommendation [4.3.2] In light of the findings of this study regarding the quality of evaluations, UNICEF should study its options in regard to the organisation of the evaluation function.

Rejected. UNICEF EO considers that this recommendation suggests a bias in the evaluator's vision of how an evaluation function should be structured, i.e. as a centralised accountability exercise. There is not sufficient comparative analysis presented to suggest that UNICEF would achieve better results in moving away from its decentralised structure. It is noted that in the 2003 ALNAP meta-evaluation covering evaluations in the humanitarian sector by UN agencies, donors, NGOs, and research institutes, there was no significant difference in overall quality of evaluations between those agencies with centralised evaluation functions (UNHCR, WFP) and UNICEF. The above-mentioned Plan of Action to Strengthen Evaluation in UNICEF was developed within the framework of the decentralised UNICEF evaluation function. It includes a full assessment of the evaluation function in 2006.

PLANNING AND STRATEGIC FOCUS

Recommendation [2.1.1] NYHQ Evaluation Office and Regional Offices should develop a three-year rolling evaluation plan, in conjunction with the country offices, to coordinate evaluation activity and ensure adequate coverage of key issues. The plan should be updated annually.

Recommendation [2.2.1] UNICEF needs some means of frequently updating country office awareness of what evaluations are being started and completed in other country offices.

Partially accepted. UNICEF is now working to develop a Global Integrated Monitoring and Evaluation Plan in the context of the forthcoming Medium Term Strategic Plan 2006-2009. This will bring together more systematically the planned evaluation work of headquarters divisions and regions.

Linked to the expansion of the Global IMEP above, Regional Offices (ROs) will establish an evaluation planning mechanism through Regional Management Teams (RMTs). This mechanism will be designed to prioritise evaluation efforts to ensure better focus on decision-makers needs. It will require a process of mapping out evaluation efforts within the region, identifying common priority areas and potential for collaboration as well as prioritising evaluation resources, not least RO or regional evaluation technical support. Different ROs and RMTs will use different processes but the intent will be the same. This regional level effort will build on and reinforce CO Integrated M&E Planning, drawing CO management attention to the focus of evaluations and rationalising the organisation's efforts within the region. Regional planning will of course need to balance country demands, including increasing joint evaluation efforts, which should be the focus of UNDAF M&E plans.

PLANNING

Recommendation [4.1.1] UNICEF should schedule evaluations of intervention strategy and performance evaluations at points where most can be learned, and/or when decisions depend on the evaluation (not only funding decisions).

Recommendation [4.3.4] Evaluations should not be rushed. Evaluations of alternative intervention strategies generally take longer than performance evaluations and should be scheduled well ahead of date at which a decision has to be made.

Accepted. Since 1993 UNICEF has required COs to carry out Integrated Monitoring and Evaluation Planning (IMEP) in conjunction with the development of Country Programmes and Annual Work Plans. One of the IMEP's functions is to help COs and their partners plan evaluations with different purpose and focus throughout the Country Programme cycle, rationalised according to key decision-makers' needs.

EVALUATION DESIGN STANDARDS AND OECD-DAC CRITERIA

Recommendation [7.2.1] UNICEF should ensure that all evaluations meet a minimum standard.

Recommendation [2.2.3] Every UNICEF evaluation should be written for a broad audience within the organisation, giving sufficient background and context to a reader unfamiliar with the country and programme, and explicitly describing the international relevance of points made in the evaluation.

Recommendation [5.2.1] All UNICEF evaluations of intervention strategy must address the wider relevance of lessons learned, and the replicability and scalability of the successful aspects of the project or programme. (See Appendix 3: Model Table of Contents for an Evaluation of Intervention Strategy.)

Recommendation [5.3.1] Every UNICEF evaluation should include a full profile of the intervention and its context, an attribution analysis where appropriate, and a consideration of replicability, scalability, sustainability, and environmental aspects.

Recommendation [6.1.3] Evaluation reports should discuss data quality and explain the effects of reach constraints, attrition, and/or non-response.

Recommendation [5.5.1] Every evaluation should include comprehensive measurement of outputs.

Recommendation [5.6.1] Not every evaluation needs to measure outcomes. In some cases, outcomes can be emphasised too much.

Recommendation [5.6.2] If outcomes are measured by the evaluators, there should be a well-based attribution analysis as well.

Recommendation [5.7.1] If the project, programme or institution is expected to be sustained, without or without continued UNICEF funding, then the mechanisms for ensuring this should be explained. If commitments from others are necessary, then these should be described. Self-sustainability of institutions and programmes through cost-recovery should be assessed where relevant.

RELATED RECOMMENDED GOOD PRACTICE

Recommendation [6.2.1] Draft findings, lessons, and recommendations should be subjected to challenge in an “exit workshop”, involving all major stakeholders.

Recommendation [4.3.5] Each evaluation should include an entry and an exit workshop with key stakeholders.

Recommendation [5.1.2] Every evaluative human-rights-based analysis should include a gender analysis.

Partially accepted. UNICEF policy on evaluation has since 2001 included reference to programme evaluation standards widely promoted in professional evaluation associations,

including the African Evaluation Association. All of the above recommendations are either covered in or follow naturally from a combination of such standards and reference to the standard OECD-DAC evaluation criteria.

UNICEF will strengthen existing references to programme evaluation standards in the 2004 revision of the Programme Policy and Procedures Manual (PPPM), making their use as a reference mandatory. References will continue to direct COs to use those standards of the national or regional professional evaluation association where they exist and, failing that, to the standards of the American or African Evaluation Associations.

Since 1991, the PPPM¹⁰⁴ has referred COs to the standard **OECD-DAC evaluation criteria** – relevance, efficiency, effectiveness, impact and sustainability. In 2002, the PPPM was revised adding reference to the recommended criteria for evaluation of humanitarian response – coordination, coherence, coverage and protection. In the 2004 PPPM, these references were revised to clarify that utilisation-focused evaluations should address standard evaluation criteria differently according to purpose of the evaluation. While this implies that not all evaluation criteria will be addressed in every evaluation, references in the PPPM now require that evaluation TORs explain both which evaluation criteria are/are not addressed and to what degree. (The existing Evaluation Technical Note no. 2 on evaluation TORS has been revised to reflect the same.) The PPPM was also revised to introduce the two new evaluation criteria related to analysis of results-based management and human rights-based approach to programming.

To strengthen gender analysis, UNICEF is developing both standards and technical guidance. The Technical Note no. 2 on evaluation TORs has been revised to highlight the importance of gender analysis. UNICEF will over the course of 2004 expand resources available on both issues on the evaluation pages of the UNICEF intranet.

Using existing material in the UNICEF M&E Training Resource, UNICEF will also highlight via intranet and the Evaluation Technical Notes series the good practice tips that help to achieve evaluation standards, including entry and exist workshops, as well as other means of increasing stakeholder participation, and gender analysis. This will be linked to work on guidance for evaluation TORs discussed below.

Most of the above issues have also been included in the Evaluation Report Standards, which are also discussed further below.

COST ANALYSIS

Recommendation [5.4.1] Every evaluation should include a cost analysis.

Recommendation [5.4.2] All performance evaluations should include an efficiency analysis - a comparison of costs and results — and all evaluations of intervention strategy should include a comparative cost analysis for each alternative being assessed.

Recommendation [5.4.3] UNICEF should have guidelines on how to estimate the full cost of a project or programme, including UNICEF staff time, contracts, costs of partners, and costs of participants.

¹⁰⁴ The PPPM was until 2000 named Book D.

Partially accepted. Cost analysis is implied in the OECD-DAC criteria of efficiency and can also be expanded in questions related to effectiveness and impact. UNICEF does seek to strengthen cost analysis in evaluation and materials providing concrete practical guidance are included in the UNICEF M&E Training Resource. These will be further profiled either on the evaluation page of the UNICEF Intranet and/or in the Evaluation Technical Notes.

However, UNICEF notes that in the context of instability and humanitarian crisis, which affects over 50 countries in which UNICEF is operating, there are significant challenges to cost analysis; even efficiency and comparative cost analyses are often prohibitively expensive. This has been well documented by Overseas Development Institute, among others. (See Hallam, 1997.) As with other OECD-DAC criteria, EO will recommend consideration and require justification as to what criteria are and are not addressed in each evaluation. However, it is neither consistent with an utilisation-focused approach nor realistic to expect cost analysis in every evaluation.

As the main concern is the lack of cost analysis in the total body of evaluation work undertaken by a given CO, in developing further guidance on IMEP, UNICEF will include reference to the importance of well-planned cost analysis of priority and high-cost programmes, projects or activities.

UNICEF-SPECIFIC EVALUATION CRITERIA — RIGHTS ANALYSIS

Recommendation [5.1.1] Every UNICEF evaluation of intervention strategy should include a human-rights-based analysis, including identifying systemic problems and potential strategies for addressing them.

Accepted. EO is developing materials on evaluation of human-rights-based programming and evaluation from a human –rights-based perspective in the course of its work on methodology development for country programme level evaluation. These will be distilled into an Evaluation Technical Note as soon as possible, at the latest 2005.

SAMPLING

Recommendation [3.3.2] The Regional Offices and/or NY HQ should provide sampling guidelines and consultation to evaluation managers in country offices.

Accepted. UNICEF considers that this recommendation is already addressed. Regional Office M&E officers and technical experts in headquarters (the Strategic Information Unit of the Division of Policy and Planning) do provide technical support to COs on issues such as sampling. General guidance is available on sampling approaches in the M&E Training Resource, including on sampling in qualitative methodologies. More detailed technical guidance on statistical sampling used in the Multi-Indicator Cluster Surveys (MICS) is included in the UNICEF End-Decade Multi-indicator Cluster Survey Manual (2000) and has been covered in depth in related training.

MANAGEMENT OF THE EVALUATION PROCESS

ETHICS

Recommendation [4.2.1] UNICEF should state an evaluation research ethics review policy, including a policy on subject competence and informed consent, and a policy on adverse event reporting.

Recommendation [4.2.2] UNICEF should institute a research ethics board, and the board should apply for accreditation, perhaps by the AAHPP.

Recommendation [4.5.1] UNICEF Evaluation Office should produce a statement of children's rights as evaluation research subjects and participants.

EO will work with DPP to establish a clear policy on research ethics in UNICEF-supported monitoring and evaluation work, specifically on UNICEF CO responsibilities. This will build on and encompass guidance already developed in the EO Technical Note no. 1 on ethical considerations in children's participation in monitoring, research, and evaluation.

However, UNICEF will not institute a research ethics board for UNICEF-supported evaluations at country level, as this is seen as an approach that could ultimately undermine national government partners. Rather, in its policy, UNICEF will recommend that UNICEF COs work with national ethics review boards where they exist and otherwise ensure ethics is included in the review of evaluation TORs proposed below.

Recommendation [4.2.3] Each evaluation report should contain a statement of how the objectivity and independence of the evaluators and their report were ensured.

Accepted. This has been integrated into the revised version of the Evaluation Technical Note No. 2 on TORs and will be considered for inclusion in the 2005 revision of the Evaluation Report Standards.

TERMS OF REFERENCE

Recommendation [3.2.3] UNICEF should develop a standard electronic form, with guidelines, for Terms of Reference for evaluations.

Accepted. In 2002, UNICEF EO issued the Evaluation Technical Note no. 2, "What goes into an Evaluation TOR", and has developed specific training materials on using the TOR as a management tool in the UNICEF M&E Training Resource. In response to the findings of the meta-evaluation, EO has already revised the Technical Note no. 2 to cover more specifically the use of evaluation criteria as appropriate to the purpose of the evaluation, management of ethical issues, and evaluation report standards (more on the latter below). In 2004/5, the EO will develop guidance further, developing a web-based tool that provides a model structure and format for evaluation TORs (as established in the Evaluation Technical Note no. 2), with links to further technical guidance and standard text options where appropriate. This work will draw on much of what is available in the M&E Training Resource. EO has also expanded its evaluation database as mentioned above to systematically include TORs for evaluation reports. This will

provide a bank of examples of TORs as models for COs, both in terms of process management and design issues.

With clear good practice standards for evaluation TORs, EO will also work with the UNICEF Office of Internal Audit (OIA) to include a systematic assessment of practice through CO audits.

Recommendation [3.3.1] The Terms of Reference for each evaluation study should state whether the study is to contain a situation analysis, an evaluation of intervention strategy and/or a performance evaluation, and for each of these that it does contain, an appropriate methodology should be described.

Recommendation [4.3.3] UNICEF evaluation managers should consult generic terms of reference when designing an evaluation, should include the topics shown in those generic TORs, and should require the evaluation table of contents to include them, unless there is good reason otherwise. (See Appendices 3 and 4.)

Partially accepted. UNICEF does not use the proposed typology of evaluation studies — situation analysis, evaluation of intervention strategy, performance evaluation — and the meta-evaluation does not present a convincing case for adopting this. As the meta-evaluation highlights, evaluations often for good reason respond to more than one purpose and many evaluations will be hybrids in terms of design. Design should be developed specifically for the evaluation.

As mentioned above, UNICEF will reinforce guidance as to the use of standard OECD-DAC evaluation criteria and more guidance will be developed as to how these are adapted.

UNICEF guidance will continue to emphasise the need for coherence between overall purpose of the evaluation, the scope, the evaluation questions and framework, and methodology. As the size of TORs on the Evaluation and Research Database grows, EO will be able to systematically identify and profile a range of model TORs to serve as a reference.

Further, as mentioned above, guidance will emphasis coherence between major evaluation questions outlined in the TORs and structure of the body of the evaluation report.

GUIDANCE ON REPORTS

Recommendation [3.2.2] UNICEF terms of reference should include a model Table of Contents for the evaluation report, such as those shown in Appendices 3 and 4.

Recommendation [6.1.2] Evaluation reports should follow a standard format, unless there is a good reason for varying it.

Partially accepted. UNICEF agrees that the majority of evaluation reports should follow a standard format. A model report structure already exists in the M&E Training Resource covering executive summary, table of contents, introduction (including independence of evaluation and methodology, limitations etc.), and the necessary appendices. The Evaluation Technical Note no. 2 on TORs does include reference to this model report structure. The Technical Note has been revised to include a draft Table of Contents among the recommended intermediate products to be submitted by evaluators for review in the course of an evaluation. Further, these same elements feature in the evaluation report standards that EO is currently

developing. However, UNICEF does not find sufficient argument in the meta-evaluation or in a brief review of the evaluations considered of excellent quality, to support standard structures for the body of evaluation reports, i.e. the evaluation findings and conclusions. Rather, the structure of evaluation findings and conclusions will typically correspond to specific evaluation frameworks and questions.

Recommendation [6.1.1] Evaluation reports should be limited to a more-or-less standard length of presentation, say 50 pages + executive summary + appendices.

Accepted. This is seen as a guide as opposed to a strict limit.

QUALITY ASSURANCE MECHANISMS AND INCENTIVES EVALUATION TORS

Recommendation [3.2.1] UNICEF should require that an electronic copy of the terms of reference should be submitted to the Region Office and NYHQ before an evaluation is contracted. This would require some adjustment to existing roles and accountabilities.

Rejected. UNICEF finds that this recommendation does not follow logically from the problem identified — that neither EO records nor the Evaluation and Research Database systematically contain TORs. Further, it does not fit with current accountability structures and RO capacity. Rather, EO has already revised the Evaluation and Research Database and the internet submission portal to systematically include TORs and track their submission. The requirement to submit evaluation TORs has been further stipulated in evaluation submissions procedures in the Country Office Annual Reporting guidelines.

EVALUATION DESIGN REVIEW

Recommendation [3.2.4] Substantial evaluations, say those budgeted at more than \$US25,000, should begin with an explicit research design exercise to produce an “evaluation framework”, and this framework should be referenced by the TORs for the subsequent evaluation.

Recommendation [3.3.2] Each evaluation study should have a methodological review and challenge by a peer in the country office, and, when the evaluation is important and difficult, by the Regional Office, or, when more appropriate, by someone in another CO. For each evaluation budgeted at over \$10,000, this review should be based on a full framework.

Partially accepted. EO will work with ROs to establish clear procedures for two levels of review of evaluation TORs, ideally in an Executive Directive to be issued in 2004. This would include that COs must establish a procedure for a peer review of all evaluation TORs, whether among peers within the CO, involving national professional evaluation associations or other appropriately qualified peer groups as the CO may choose. Further, it would include stipulation that over a specified UNICEF financial contribution, e.g. US\$25,000, and/or for evaluations identified within the Regional Management Team mechanism as major evaluations, evaluation TORs must be reviewed by either the RO or an RMT-established peer review mechanism. Both levels of review must include a technical review of evaluation design (e.g. evaluation frameworks and methodology). Once established, EO would work with OIA to include CO review of evaluation TORs among the issues covered in CO audits.

Recommendation [4.3.7] To ensure minimum quality standards are met, appropriate management, evaluation and sector expert sign-offs should be mandatory, particularly at the research design stage of the evaluation.

Rejected. As mentioned above, UNICEF will rather make this an issue for review in the course of CO audits.

FOLLOW-UP TO EVALUATIONS

Recommendation [6.3.1] The UNICEF country office should record its decision on each evaluation recommendation.

Recommendation [6.3.2] For each evaluation plan that contains recommendations that the country office accepts, the evaluation manager should prepare an action plan for approval by the CO senior management.

Recommendation [6.3.3] The approved action plan should become an appendix to the evaluation report. – THIS DOES NOT EXIST....

Partially accepted. At the global level, the UNICEF Evaluation Committee, constituted in August 2003, now reviews evaluation reports that have relevance at the global governance level. In this review process, the Committee reviews the evaluation reports, the management's responses and proposed action plans for the implementation of the recommendations. The action plans for these evaluations will be recorded in the Evaluation Database.

At regional level, RMTs will be establishing a mechanism for reviewing key evaluations and formulating response to issues for regional governance. The functioning would be comparable to the global UNICEF Evaluation Committee. Attention to evaluation results among CO and RO senior managers would support CO management attention to the design and product. It would also contribute to learning within the region.

To further reinforce follow-up of the remaining majority of programme evaluations at the country level, explicit responsibilities for follow-up procedures have been established in the 2004 UNICEF PPPM. CO are made responsible for ensuring that results of all evaluations are explicitly discussed and follow-up actions recorded in a meeting of the UNICEF Country Management Team and/or another appropriate accountable body, internal to the CO and/or with partners as fits with the nature of the recommendations.

For UNICEF, it is important to see this accountability in the context of a more general commitment by UNICEF to the principle of involving key stakeholders in the evaluation process to heighten ownership of evaluation results and commitment to follow-up. Existing guidance in UNICEF M&E Training Resource emphasises involvement of key stakeholders at least in the development of the evaluation design and in the validation and analysis of findings.

EO will work with OIA to introduce a more systematic check on this process in CO audits. EO will also explore with ROs and the Division of Information Technology, both the demand for and feasibility of establishing a global electronic database to facilitate COs recording follow-up to evaluations. Options of linking such records to UNICEF's Programme manager system (ProMS) or to UNICEF evaluation databases will be explored.

Recommendation [6.3.3] Where an evaluation recommendation requires action by another agency, UNICEF should ask that agency for a response. If the response is positive, the agency should be asked for an action plan. If an action plan is forthcoming, it should be appended to the evaluation report.

Rejected. This is considered far too unwieldy given current capacities.

OTHER PUBLIC ACCESS TO EVALUATIONS

Recommendation [2.2.4] To facilitate openness and to give an incentive to improve quality, all UNICEF evaluations that cost more than \$10,000 should be available to the public on the open website.

Partially accepted. The vast majority of evaluations submitted to the EO are accessible to the public on the Internet. The UNICEF Evaluation and Research Database was accessible on the UNICEF Intranet from 2001 and on the Internet from 2002. However, the ERD was explicitly revamped in 2000 as a learning tool. To improve the ERD as a learning tool, EO has developed and issued Evaluation Report Standards. EO will use these standards to select those reports considered adequate for inclusion in the Evaluation and Research Database. Similarly, using the ratings against standards, UNICEF will be able to systematically identify, profile, and disseminate “good” evaluation design.

UNICEF does recognise that low CO compliance with requirements to submit evaluation reports to EO are a problem, not least because this reduces the total pool from which more good evaluations could be drawn for posting on the ERD. As mentioned above, EO has just recently launched a new evaluation report submission portal designed to make it easier technologically for COs to comply as well as to allow EO, ROs and COs to track report submission.

EO will seek to clarify the organisation’s disclosure policy in 2004. The proposed disclosure policy will: establish that evaluation reports should in principle be made public; allow exceptions where the treatment of politically sensitive issues is expected to cause undue damage to an organisation/institution; establish clear criteria for any such exceptions; establish the responsibility of COs, regional or headquarters offices to specify in writing to the Director of Evaluation any evaluations that they have supported that should not be disclosed to the public; and re-establish the responsibility of all UNICEF offices to submit reports for all the evaluations they have supported to the Evaluation Office for inclusion in the corporate evaluation database. The latter refers to a document repository that is maintained for institutional memory and allows the organisation to verify both the quality of evaluations and decisions on disclosure.

In terms of addressing incentives to good quality evaluation, internal analysis suggests a number of more promising avenues that the EO and ROs have committed to pursue. As mentioned above, these include using the forum of the Regional Management Team and working with OIA to introduce the review of key aspects of evaluation management in CO audits (i.e. use of IMEP, use of TORs and programme evaluation standards, quality review of TORs, and follow-up to evaluation).

Recommendation [2.2.2] The full text of completed evaluations from the current year, and two previous years, should be accessible on the Internet in Adobe .pdf format, and key worded for easy search and access. (Linked in the meta-evaluation to a reported perception by CO staff of difficulties in accessing evaluation reports.)

Partially accepted. As mentioned above, the ERD was accessible on the UNICEF intranet from 2001 and the Internet from 2002. EO has made significant improvements in the design, including attachment of full reports and now TORs in pdf. An ongoing corporate upgrade to the UNICEF Internet site and subsequent migration of data should allow better search capability by 2005.

UNICEF considers that establishing a user-friendly database on the Internet is not the key challenge in accessibility of evaluations for COs. Rather, EO is focusing on ensuring quality content, stimulating the proportion of evaluation reports actually submitted to the ERD, and increasing awareness of the ERD.

In terms of quality of content, EO has increased support in headquarters to ensure quality executive summaries are posted, has introduced Evaluation Report Standards to ensure more consistent quality of the actual reports posted and will expand its efforts to profile and disseminate high quality evaluations in 2004. As part of this effort, EO will carry out a client survey in 2005 to more effectively define the best approaches to bring lessons from evaluation back to decision-makers in the field, as well as in regional and headquarters offices. This will in turn increase visibility of the ERD.

As mentioned above, EO will pursue a number of initiatives to increase submissions to the ERD. Finally, EO notes that the issue of slow or otherwise limited Internet access for COs is dramatically changing as country infrastructure develops, but remains an issue.

LANGUAGE OF DISSEMINATION

Recommendation [4.3.6] An expanded executive summary of the evaluation report should be translated into the local language[s]. This should be part of the Terms of Reference and appropriately budgeted.

Accepted. The CO responsibility to disseminate evaluation results to key stakeholders, including international and national partners, is established in the UNICEF PPPM. EO will make explicit CO responsibilities to translate evaluation executive summaries and/or other summary pieces tailored to key national audiences.

The EO notes that the meta-evaluation sample is biased toward reports available in English. CO reporting on evaluations undertaken suggests that many evaluation reports are produced only in the national language and not translated. Given the limited resources for translation of complete reports, it has been agreed that the institutional database EO maintains for all evaluations shall include reports in all languages. The Evaluation and Research Database on UNICEF intranet and Internet will maintain an English executive summary of all evaluations that meet the quality criteria that will be developed. Spanish and French-language executive summaries will be added to the Internet site when it can accommodate this.

CAPACITY BUILDING

Recommendation [4.4.1] UNICEF should commission an evaluation of alternative ways to improve national evaluation capacity.

Accepted. The Plan of Action to Strengthen Evaluation in UNICEF includes such an evaluation to be undertaken in 2005.

CONTRACTING WITH “IMPLEMENTING AGENTS”

Recommendation [5.5.2] UNICEF should consider making contracts/agreements with implementing agents more output-performance-based in order to facilitate results-based management and to enable good performance evaluation.

Rejected. UNICEF finds this recommendation is not substantiated by the findings. Further, contracting is not convincingly argued as the highest priority entry point for strengthening results-based management with UNICEF partners. UNICEF is working to strengthen results-based programme planning and management, with emphasis on a range of tools from macro five-year results matrices to the above-mentioned IMEP and annual management tools.

BUDGET ALLOCATED

Recommendation [4.3.1] UNICEF should undertake a study of the costs of evaluations, and produce guidelines for appropriate budgeting of situation analyses, evaluations of intervention strategy, and performance evaluations. The option of fewer but better studies should be assessed. Evaluation budgets should be adequate to allow a thorough evaluation design, sufficient time in fieldwork to obtain good data, and adequate time for analysis and reflection.

Rejected. EO recommends that such a study would be extremely complex given the range of country and logistical contexts within which UNICEF-supported evaluations are undertaken, as well as the range of typologies of evaluation across different programming areas. Further, analysis by EO and ROs of factors influencing resource allocation to evaluation do not suggest that benchmark figures would significantly change practice. Rather, EO and ROs will continue to pursue strategies mentioned above to promote and strengthen strategic utilisation-focused evaluation.

