A Metaevaluation of NGO Evaluations conducted under the AusAID NGO Cooperation Program

November 2006

Colin Reynolds
Contact: colin.reynolds@couragepartners.com.au

Acknowledgement:

This metaevaluation relates to a sample of twenty evaluation documents drawn from twelve ANCP NGOs. It attempts to examine and identify learning opportunities for the NGO Sector. To do this it has drawn real examples from the sample documents to illustrate specific points made. Given that any evaluation must critically review some subject matter in order to suggest improvements, a number of the examples used are unavoidably presented in a negative context. It is stressed that these examples refer only to specific aspects of documents and do not imply serious flaws with any particular evaluation in its entirety.

AusAID and the author wish to offer our sincere gratitude to the twelve contributing organisations (listed in Annex 1) for their willingness to participate in this metaevaluation, and reassure them that regard for this willingness far outweighs any of the relatively minor criticisms referred to in the report itself.

Common Abbreviations Used:

ABM	**Anglican Board of Mission**
ACFID:	**Australian Council For International Development**
ADPlan:	**ANCP Annual Development Plan**
ALWS:	**Australian Lutheran World Service**
ANCP:	**AusAID NGO Cooperation Program**
APHEDA:	**Union Aid Abroad - Australian People for Health, Education and Development Abroad**
ARC:	**Australian Red Cross**
AusAID:	**Australian Agency for International Development**
DAC:	**Development Assistance Committee of the OECD**
ECCD:	**Early childhood care and development**
HIV/AIDS:	**Human Immunodeficiency Virus / Acquired Immune Deficiency Syndrome**
IPF	**Indicative Planning Figure**
IWDA:	**International Women's Development Agency**
LWF:	**Lutheran World Federation**
M&E:	**Monitoring and Evaluation**
NGO:	**Non-Government Organisation**
OECD:	**Organisation for Economic Cooperation and Development**
PDR:	**People's Democratic Republic**
PLWHA:	**People Living with HIV/AIDS**
PNG:	**Papua New Guinea**
RDE:	**Recognised Development Expenditure**
SAADO:	**Salvation Army of Australia Development Office**
SCA:	**Save the Children Australia**
SH&FPA:	**Sexual Health & Family Planning Australia**
ToR:	**Terms of Reference**
UCOA:	**United Church Overseas Aid**
UNICEF:	**United Nations Children's Emergency Fund**
WHO:	**World Health Organisation**
WVA:	**World Vision Australia**

Table of Contents:

EXECUTIVE SUMMARY

A. Approach

The Terms of Reference (TOR) of this metaevaluation stated the following objectives:

  A. Determine the quality of a cross-section of ANCP activity evaluations;

  B. Determine the validity of a cross-section of ANCP activity evaluations' results and recommendations; and

  C. Determine how broadly and consistently evaluation funds are being utilized by accredited NGOs.

The application of the concept of 'quality' in the first objective was expanded to allow joint consideration of Objectives A & B. A reading of 20 sample evaluations was made and 8 key quality issues were identified (see below). These issues were then discussed separately using illustrative examples drawn from the sample.

Objective C was attempted using the information provided by ADPlans (and associated file items) over the financial years 2004/5-2006/7.

B. Findings

1. Evaluation Objective A (Determine the quality of a cross-section of ANCP activity evaluations) & Evaluation Objective B (Determine the validity of a cross-section of ANCP activity evaluations' results and recommendations)

These objectives were considered jointly by the metaevaluation. In regard to the eight key quality issues identified in the sample evaluations, the following (rough) proportions of reports displayed significant **shortcomings**:

  1. Level to which the evaluation provides a balance of *Accountability* and *Continuous Improvement* Considerations – **one quarter**;
  2. Level to which assertions made (especially final conclusions and recommendations), are supported in a defensible and methodologically sound manner – **just under half**;
  3. The drawing out of 'lessons learnt' and their usefulness - **just under half**;
  4. Clarity and conciseness in presentation - **just under half**;
  5. Independence of evaluators – **one quarter**;
  6. Level to which project M&E issues are raised, particularly in regard to *Risk Management* - **just under half for M&E (only one tenth addressed Risk)**;
  7. Attempting to move beyond 'output-level' analysis, to consider modest 'outcomes' – **just over half**; and
  8. Use of participatory techniques **– almost all of the evaluations that substantially used participatory techniques displayed shortcomings in their analysis.**

Most of these shortcomings are very basic, and therefore easily addressed by making future evaluators aware of the potential pitfalls related to each issue. The sample evaluations provided valuable illustrative examples.

Recommendation 1. Organisations receiving ANCP subsidies be required to provide a copy of this report to anyone tasked with organising or carrying out future evaluative exercises; including both internal staff and external contractors.

Recommendation 2. ACFID (Australian Council For International Development) be asked to provide copy of this report to anyone commissioned by them to provide collective M&E training to the NGO community, so that it may be used to help prioritise content of such training.

Recommendation 3. AusAID include in the TOR of the next ANCP metaevaluation a requirement to consider whether improvement has been made in regard the 8 key quality issues identified in this report.

2). Evaluation Objective C (Determine how broadly and consistently evaluation funds are being utilised by accredited NGOs)

This objective was attempted in relation to the financial years 2004/5-2006/7 of ANCP subsidies, but proved difficult to reliably fulfill given inconsistencies in reporting. In broad terms, it appeared from the available data that approximately 30% of ANCP NGOs were not undertaking evaluative activities of any kind. Constraints were identified in regard to the shared understanding of the term 'evaluation' and hence, what may be funded using the ANCP 5% allowance. Some ANCP NGOs appeared to display lack of balance between smaller and more regular formative evaluative exercises and 'occasional' major summative evaluations, with a bias toward the latter. It was noted that the ANCP 5% allowance alone was not sufficient to encourage NGOs receiving a low level of subsidy to maintain an appropriate evaluation regime, and that associated compromises made were often counter productive.

Recommendation 4. AusAID clarify what it means by 'evaluation', particularly in regard to the distinction between 'monitoring' and 'evaluation' (consideration may be given to using the specific suggestions made in section 5.3.1).

Recommendation 5. AusAID require that the names of all reports produced through evaluation expenditure be provided in 'Significant Outputs' sections of ADPlan Reports.

Recommendation 6. AusAID consider simplifying the process for rolling over unused evaluation funds to future years (including options for NGOs receiving greater than $200 000 p.a.), and establishing a modest ($50,000-$100,000) 'major evaluation fund' (external to ADPlans), from which NGOs receiving a low level of ANCP subsidy can occasionally draw to conduct major evaluative exercises.
.

1. BACKGROUND

The AusAID NGO Cooperation Program (ANCP) is an annual funding program for Australian NGOs that have been accredited with AusAID. The goal of the ANCP is to subsidise Australian NGO community development activities which directly and tangibly alleviate poverty in developing countries. Under ANCP guidelines, recognising that evaluation is a normal part of the project management cycle, NGOs can utilise up to 5% of their annually allocated ANCP funding for evaluations.

In 2004 AusAID and the Committee on Development Cooperation (CDC) were presented with a "Review of the AusAID NGO Accreditation Process and Systems". In this review it was recommended that "AusAID gather annual evaluation reports from agencies (funded by the 5% that ANCP allows for Evaluation) and undertake a meta-evaluation."

With the development of new initiatives under the *White Paper On The Australian Government's Overseas Aid Program* (2006), it is now timely to undertake a limited evaluation of the ANCP funded evaluations and relevant documents to determine their quality and validity. It is also timely for AusAID to examine the broader context around evaluations in order to inform future ANCP directions. In this case, this process is termed a "metaevaluation".

2. INTRODUCTION

The task of evaluating other evaluations is always a contentious one. There are many equally valid approaches and methods which may be applied in evaluations, so it would be inappropriate for a metaevaluation to simply look for favoured options. This is particularly so in this case, given the wide variety in the sample of evaluations considered. However, this does not prevent the proposing of a basic minimum standard.

For the purposes of this metaevaluation, the minimum standard is based on the assumption that evaluative activities are intended to serve at least two purposes. The first is the provision of *Accountability;* ensuring[1] that implementation was sincerely attempted. The second purpose is to provide the basis of *Continuous Improvement* in interventions; whether in a *formative* (feeding into the current project cycle) or *summative* (feeding into future project cycles) manner[2].

> **Note, this view is generally in line with the ACFID Code of Conduct Guidance Document, particularly Appendix 7 (NGO Effectiveness Framework –June 2004) which includes the following:**
>
> **"Given the dynamic nature of development, Australian NGOs and their implementing partners seek to undertake ongoing reflection about their work and the context of that work, for the purpose of improvement" (p.2);**

---

[1] And reassuring donors.

[2] In practice, most M&E exercises cover both formative and summative functions. While purely formative 'evaluations' (e.g. feasibility studies or design missions) do not require an accountability role, no purely formative exercises were included in the sample. For more comprehensive definitions of 'Accountability', 'Formative Evaluation' and 'Summative Evaluation' see: http://www.oecd.org/dataoecd/29/21/2754804.pdf

"Australian NGOs, together with their implementing partners undertake monitoring and evaluation of their work, in order to identify areas of success, areas requiring improvement and in order to implement timely and appropriate change" (p.3); and

"Australian NGOs work in ways that include clear and transparent accountability mechanisms to all key stakeholders" (p.3).

3. METHOD

The terms of reference (TOR) to this exercise are attached as Annex 3. The objectives of this desk-based metaevaluation of ANCP funded evaluations were to:

A. Determine the quality of a cross-section of ANCP activity evaluations;

B. Determine the validity of a cross-section of ANCP activity evaluations' results and recommendations; and

C. Determine how broadly and consistently evaluation funds are being utilized by accredited NGOs.

3.1 Evaluation Objectives A & B

3.1.1 Evaluation 'Quality'

For the purposes of this metaevaluation, the concept of 'quality' has been broadened to allow joint consideration the first two objectives. While Objective A is clearly a core element of any metaevaluation, it is only feasible for a deskbound study to verify if evaluation findings are well-argued and defensible, rather than a true representation of conditions on-the-ground. This is consistent with the TOR which unpacks Objective B into:

- An examination of the findings of the evaluations and if they are defensible; and
- An examination of lessons learnt.

A clear implication of this approach is that any praise or criticism provided in relation to an evaluation does not necessarily translate into praise or criticism of the project being evaluated. Indeed, it was not the intent of this metaevaluation to attempt to *re-evaluate* the projects themselves.

3.1.2 The Sample

The twenty evaluations under scrutiny vary greatly. Some are largely summative completion or ex-post analyses feeding into the next cycle of project design. Others are largely formative reviews intended to improve the current project cycle. Some are the work of sizable and highly qualified teams, inclusive of independent members. Others are the work of one person from within the agency responsible for project implementation. Some have consumed considerable financial resources. Others have been very constrained in this regard.

Many of these differences have been based on justifiably pragmatic decisions, given the desired outcomes of the evaluations and the relative 'sizes' of the projects themselves; both in regard to total cost and the level of on-the-ground resources available to conduct studies. It is therefore inappropriate to review these evaluations against some single, abstract 'ideal'[3]. Like the projects they address, no evaluation is perfect. They all consist of a range of aspects, some of which may be done well and others less so. In the absence of a set of idealized criteria, the metaevaluator has simply attempted to review each document from the perspective of an interested reader. **A reactive, rather than preemptive definition of 'quality' has thereby been applied, allowing the sample itself to suggest the most pertinent quality issues.**

While this approach unavoidably involved some level of subjectivity in regard to the assumed levels of technical knowledge, familiarity with projects, and time available to scrutinise reports, this is largely overcome by consistently favouring the lower end of such scales. In hindsight, this appears to have been a fairly useful position to take, because many of the issues identified in reports (beyond the minimum standard proposed above), are very basic and often relate to report writing in general, rather than specifically to evaluations. In short, the sample suggests that a 'back to basics' focus is likely to be most productive for this document.

Given that evaluations are multi-faceted, and that the emphasis of each may legitimately vary with the type of evaluative exercise being undertaken, it would also make little sense to try to 'rank' the twenty sample evaluations from 'best to worst'[4]. This metaevaluation was not regarded as a 'competition' between those in the sample. Rather, it was used as an opportunity to provide **practical explanations and illustrative advice** regarding the most common issues identified across the board.

3.1.3 Defining "Quality" in the Context of the Sample: Identifying Key Quality Issues

An initial reading of the sample evaluations was made to draw out common strengths and weaknesses. While some reliance is admittedly placed on the 20 examples being a representative sample, it is hoped that this approach will be sufficiently comprehensive to address the currently pertinent quality aspects of most ANCP evaluations[5]. From the initial reading, eight key quality issues were identified as useful discussion topics for addressing the quality of the sample as a whole.

These **key quality issues** are:

1. Level to which the evaluation provides a balance of *Accountability* and *Continuous Improvement* considerations;

---

[3] Such 'best-practice example' approaches also often encourage only contextually-blind and overly formulaic mimicry, rather than facilitate more practical continuous improvement measures.
[4] Again, this would amount to making an absolute comparison of unlike entities.
[5] Note that this limitation specifically avoids implying that the aspects covered are comprehensive in regard to evaluations beyond the ANCP context.

2. Level to which assertions made (especially final conclusions and recommendations) are supported in a defensible and methodologically sound manner;
3. The drawing out of 'lessons learnt' and their usefulness;
4. Clarity and conciseness in presentation[6];
5. Independence of evaluators;
6. Level to which project M&E issues are raised, particularly in regard to *Risk Management*;
7. Attempting to move beyond output-level analysis, to consider modest 'outcomes';
8. Use of participatory techniques.

Hence, for the purposes of this metaevaluation, these 8 issues form the criteria against which the 'quality' of the set of sample evaluations is considered. As they are simply a basis for discussion, these issues are not intended to be mutually exclusive. Each issue has been addressed separately (Section 4), with associated discussions attempting to illustrate 'lessons' for future ANCP evaluations.

3.2 Objective C

Data collection for this part of the report consisted of examining three years (2004-07) of relevant ADPlans (and, where possible, associated file entries) looking for evidence of funds being used for evaluative purposes. It was soon discovered that the financial and other data provided in these documents was not consistent enough to identify detailed trends. However, it was sufficient to raise many potential concerns. Quantitative analysis of this data is therefore limited to broad overviews. More detailed qualitative analysis of entries which raise concerns has also been provided.

4. KEY QUALITY ISSUES OF ANCP EVALUATIONS

4.1 Level to which the evaluation provides a balance of *Accountability* and *Continuous Improvement* considerations

As already discussed above (Section 2.), the fulfilling of these two purposes is regarded as a minimum standard of evaluations. This was one of the better performing considerations, with around three quarters[7] of the sample achieving this minimum standard.

---

[6] Note that in deference to ESL contributors, this did not include any grammatical considerations whatsoever. While clear arguments will always shine through poor English, the best English cannot aid a confused argument.

[7] Note that as all of these key issues can be assessed in terms of 'degree of applicability', this figure **and all other proportions provided elsewhere in this section** are necessarily subjective and therefore should be considered only roughly indicative. While less defensible, this approach avoids the need to separate out reports of particular ANCP NGOs in each case. To do so would be inappropriate, as many reports were produced by external evaluators.

4.1 Example A.
From: Evaluation of Four Projects on the Thai-Burma Border (August 2003)

This report provides the objective of the evaluation as:

  "The objective of the evaluation was to review IWDA support of four ethnic community organisations on the Thai-Burma border:

  1. **To provide baseline information on which to measure changes over the next year.**
  2. **To assess the partner agency links with the community and with other agencies.**
  3. **To explore what is happening for the organisation and other women's ethnic organisations on the Thai-Burma border so that IWDA can link its work and advocacy more effectively with that of other partners in the region.**
  4. **To identify other ways in which we can assist the partner in future.**
  5. **To meet with staff in the partner organisation and strengthen the working relationship between IWDA and partner agency." (p.4)[8]**

While all potential aspects of an evaluation, none of these directly address progress made to date in relation to the projects objectives[9], and hence the evaluation's accountability role is not met. Although the eventual content of the report is not entirely devoid of accountability issues, its structure is largely dictated by the above 5 sub-objectives and therefore such considerations are significantly impeded.

4.1 Example B.
From: China Integrated Development Programme Evaluation Report (March 2005)
India Health Appraisal and Rural Empowerment Project Evaluation Report (November 2004)
Indonesia Papua Health and Education Programme Evaluation Report (March 2005)

These three (related) reports appear skewed towards the accountability purpose of the evaluation. They almost exclusively focus on reporting whether 'targets were met or not'. Given that the projects addressed by the first two reports (China and India) met or exceeded their targets, these read largely as a 'justification' of the interventions. What is missing is significant analysis into why such apparent success was attained, if it could have been done 'even better', and whether it was conducive to desirable impacts. The sincerity of the evaluators is not questioned, as in the third evaluation (Indonesia) they meet a project that was much further from perfection, and they duly reported the failures to meet targets. Again, however, the 'why' of the situation received insufficient attention. This is most easily illustrated using examples of the 'lessons' that were provided in these reports[10]:

  "Lesson: Exploration and analysis is required to learn lessons as to how to advance from pilot projects." (China p 4)

  "Lesson: Good management systems and professional expertise ensure good implementation." (China p. 4)

  "Lesson: Well designed and carefully managed projects achieve targets." (India p.5)

  "Lesson: Projects should be designed with sustainability in mind, taking careful consideration of the resources and capacity required and available in the long term." (India p.6)

  "Lesson: Good project management and accountability requires coherent and transparent [financial] systems." (Indonesia p.5)

  "Lesson: Water and sanitation projects require technical expertise". (Indonesia p.5)

Lessons of this form tend to suggest that any examination of contextual, 'causal' factors was limited.

---

[8] TOR not provided for comparison.
[9] Although '2.' may have partial relevance.

4.2 Level to which assertions made (especially final conclusions and recommendations), are supported in a defensible and methodologically sound manner

The aim of any analysis is to *convince* the reader of a set of claims or propositions. In the case of evaluative documents, this is important from both donor and implementer perspectives. Both need to feel secure that any (positive or negative) claims made in regard to a project are valid, so they can undertake informed future decision-making.

Common **lessons** arising from the sample in relation to this point were:

Lesson 4.2a: It is not sufficient for an evaluation to simply 'say' that the project has done well or poorly in any particular respect, there is a need to provide clear and impartial evidence of such success or failure.

An evaluation which falls into this trap will generally read as either an attempt to a 'justify or vilify' a project[11]. Provision of some analytical content (leading to claims) is essential, and care must be taken to ensure that such analytic content does not (actively or passively) display bias.

One important part of presenting arguments in an impartial way is to ensure that a 'neutral voice' is maintained in all commentary[12]. This means carefully avoiding emotive or obsequious language (excludes quotations).

While no evaluation is likely to be perfect in this regard, at least half of the sample was regarded as having included too much unsubstantiated content.

> **4.2a Example A**
> **From: Project/Program Evaluation, Montanosa Potable Water Supply Project/Cordillera Water Program, Philippines (July 2005)**
>
> This is a five page report to address two fairly complex projects, so it is not surprising that it reads as an executive summary. It is largely subjective and contains little analysis per se, rather it makes claims that the reader can neither confirm nor deny. In fairness to the partner organisation involved, it is recognised that donors sometimes over-enthusiastically encourage brevity.   While conciseness is a virtue in many respects, even this has its limits.
>
> **4.2a Example B**
> **From: Evaluation Report, Women's Action For Change Sexual Minorities Project, Fiji (Sept. 2003)**
>
> While analyses of verifiable qualitative information are equally as defensible as those using verifiable quantitative data, the former are not always governed by as well-defined (eg. mathematical), rules[13]. Maintaining objectivity in a qualitative analysis therefore requires more care. In this document, the analysis of interviews consists solely of selected quotations from each of the interviewees. No qualification of participant perspective (see 4.8 below), nor summary consideration of views was provided. While not suggesting it has happened in this case, it is possible to take quotations out of context and be selective in what is quoted. Either of these scenarios amounts to introducing a high degree of researcher subjectivity into an analysis. Whether the qualitative data is

---

[10] Lessons usually being the result of continuous improvement analyses.

[11] In both cases the reader is likely to conclude more about the author than they may about the project in question.

[12] Note this does not merely equate 'passive' tenses.

[13] Not that these are always followed (see **4.2e Examples A** & **B**).

in the form of quotations or something else, evaluators must, in the interests of defensibility, attempt to demonstrate that they have not been biased in inclusions.

4.2a Example C
From: Faith Mission Health and Education Services Review (July 2005)

This example relates to a language issue. More specifically, it is about maintaining a 'neutral tone' and avoiding emotive or 'overly-flattering' language in the text of professional reports (excepting quotations). To some extent, this may be a regarded as 'cultural' debate between authors, but anyone writing a 'professional' report needs to remain aware of the associated standards expected by readers.   Emotive or overly flattering language tends to give a reader the impression that an author is attempting to press a point beyond the evidence supplied by supporting information. It again gives an (in this case undeserved), impression of bias. For example:

  "There is an air of limitless gratitude on the part of FMHE [Faith Mission Health and Education Services] towards SCiPNG [Save the Children Fund PNG] for all it had done". (p.9)

  "FMHE continues to respond to the requests from the communities in their area with immense care and consideration of the needs that were being identified" (p.10)

More extreme examples are available from the report text, but these two sufficiently illustrate the point. The second example is provided as a 'benchmark' of sorts, because it is a borderline case. If the adjective 'immense' is removed, or is substituted with one a little more constrained (e.g. 'great'), it becomes a very reasonable finding. It is recognised that it may be very difficult for those with English as a second language to distinguish between apparently equivalent synonyms. The intention here is not to criticise, but bring a potential pitfall to people's attention.

Lesson 4.2b: *It is not appropriate to simply refer the reader to annexes, of raw data or partial analyses, as evidence of claims made.*

It is the job of the evaluator to conduct and present analyses (albeit in summary form). Deferring overall analyses to the reader calls into question the logic of providing *any* narrative lead-in to such annexes, beyond being a cursory attempt to 'lead' interpretation of the information so supplied.

This was one of the least encountered problems, with less than a quarter of reports displaying it to any substantial degree.

4.2b Example A
From: Evaluation of Early Childhood Care and Development Project – India (May 2005)

Once read carefully from cover to cover (including annexes), the findings of this report are generally supported. However, if only the main body of the report is assessed, they appear largely unsubstantiated. The main body reads more as an extended executive summary, than the core discussion of a report. Important partial analyses contained in annexes are not referred to in any detail, nor considered as a whole. Given that many readers will tend to give far less attention to annexes, the key elements of supportive analyses need to brought forward into key discussions, otherwise such findings and the recommendations they generate will have an unsatisfying 'out-of-the-blue' quality to many readers. This is not aided by over generalised statements of findings. For example:

  "Technical know how of staff has gone up; so has the understanding of ECCD [Early Childhood Care and Development] and it's thematic relevance to other initiatives…" (p.10)

Many other reports in the sample also share this problem to varying extents, so this example is not an isolated case. A partial exception to this observation is that the main body of this report did contain a number of boxes considering the 'counter-factual' standpoint (i.e. what would the situation have been in the absence of the project). This is one of the few reports that attempted this useful approach.

4.2b Example B
From: Evaluation of Early Childhood Care and Development Project – Vietnam (April 2005)

This report provides a valuable contrast to the above example, in that its incorporation of analyses and their relation to results could not be made more explicit in the main body. It contains a complete section entitled "4. RESULTS AND ANALYSES" (p.13). While it is not suggested that this should be adopted as a 'standard' inclusion (other document structures may equally achieve the same purpose), it does prompt an appropriate rigour.

Lesson 4.2c: *Recommendations must be sufficiently detailed to be useful and be supported by relevant dual-level analyses.*

Vague or overly general recommendations are often of little practical guidance within a specific context. In addition, recommendations require two levels of analytic support. The first, convincing the reader that action/change of some form is required. The second, explaining why the *suggested* course of action is likely to be the most appropriate in the circumstances. Recommendations which appear 'out-of-the-blue', in relation to either of these two considerations are a fairly clear warning-sign that an arbitrary set of 'good ideas' has been solicited and subsequently presented in an uncritical manner (see also 4.8). Note that a key part of the second level of analytic support necessarily includes justification of the *practicality* of the suggested action[14].

While there were isolated exceptions within documents, virtually no reports consistently maintained rigour in respect to justifying recommendations on both levels.

4.2c Example A
From: Report on the Impact and Degree of Success that ARC and Vietnam Red Cross (VNRC) can Claim in Relation to the Achievement of the HIV/AIDS Program Objectives (November 2004)

While the analysis of problems addressed by this report is sufficiently critical and well argued, it tends not to explore potential solutions to any level of detail. Consequently, the majority of recommendations made tend to be high-level statements that may provide limited practical instruction to those wishing to implement them. For example:

- **"Establish a new sense of trust between the VNRC leadership and ARC Technical Assistance";**
- **"To develop a meaningful sense of ownership for the work amongst all levels of the VNRC, including the Headquarters";**
- **"To have a well-structured and systematic capacity building plan"; and**
- **"To have a coherent in-county strategy guiding the Program work (that can be used as a road-map for developing the Program)." (all p.21)**

This is a common failing of the sample evaluations, and all authors should review their own recommendations to see whether they are best described as 'practical guidance' or merely 'noble desires'.

---

[14] It is unhelpful to simply suggest desired 'outcomes', and leave the implementers to try to work out how (or if) they may be achieved. The same applies to suggestions for M&E 'indicators' (see Section 4.6) in that *how* something is to be practically measured must be addressed.

4.2c Example B
From: Evaluation of the Mongolian Red Cross Society (MRCS) Strengthening Public Health (HIV/AIDS and Blood Safety) (July 2005)

While not a consistent trait, and not always supported by sufficient analysis, some of the recommendations included in this report provide a good example of addressing a useful level of contextual detail. Examples include:

  " Initiate discussions with UNICEF, The Ministry of Education [MoE] and WHO re school based peer education strategies, including capacity building of MoE, development of national curriculum and in service teacher training. While targeting should be based on vulnerability, if MRCS [Mongolian Red Cross Society] continues to target in-school youth, it is essential that schools be selected in locations with highest HIV Vulnerability." (p.8)

  "MRCS could adapt stigma and discrimination training materials in use in other countries to their situation and work with PLWHA [People Living With HIV/AIDS] from 'Positive Life' to conduct training for MRCS staff at headquarters, and branch staff and volunteers engaged in HIV/AIDS and the Safer Blood projects in the first instance." (p.11)

  "MRCS could capitalise on the relationship the Safer Blood project is developing with monks in order to involve their leaders in promoting non-discrimination and tolerance towards vulnerable marginalised people, including PLWHA and their families." (p.11)

It is of value to contrast these examples with the recommendations quoted in relation to the proceeding example. Note also that, while the report from which the proceeding example is drawn is targeted at the program level, this should just mean a change to the context of the detail provided in recommendations, not a reduction of detail per se.

Lesson 4.2d: While supporting analyses may be based on qualitative or quantitative information, they need to be presented in a structured manner, with at least part of this structure directly focused upon the stated objectives of the project.

As a basic standard, those objectives explicitly stated in project documentation need to be substantially referred to in the evaluation. This basis may be built on by the added consideration of higher-level objectives drawn from either related documents[15] or an abstracted list of 'ideals' (e.g. the DAC[16] evaluative criteria of Relevance, Effectiveness, Efficiency, etc.). However, it should be noted that whenever such higher-level objectives are used as an overlay to analyses, they must eventually be related back to the stated project objectives (e.g. the DAC analysis should consist of considering the Relevance, Effectiveness, Efficiency, etc. of project *objectives*). Purely abstract application of 'ideal' evaluation criteria neglects the often subtle, *contextual purpose[17]* of intervention design and is not recommended.

At least half of the reports in the sample could have significantly improved the structure of their (overall) analyses.

---

[15] E.g. the objectives of the evaluation, a program strategy, or an overarching gender policy.
[16] Development Assistance Committee of the OECD (Organisation for Economic Co-operation and Development)
[17] E.g. a food for work program may not be the most efficient means of building a road in absolute terms, but it still may be the best contextual option, given other project objectives or cultural sensitivities.

4.2d Example A
From: Evaluation of Early Childhood Care and Development Project – Vietnam (April 2005)

This report provides a good example of maintaining an appropriate focus on project objectives, even when an abstract overlay of 'higher-level' objectives is also applied. The overlay used has five elements "Relevance, Efficiency, Effectiveness, Impact and Sustainability" (p.4). As can particularly be seen in Section 4.2 of the report discussing the 'Effectiveness' criterion (p.14), original project objectives are explicitly referenced and addressed during associated discussions (e.g. under "4.21 Access Objectives" [p.14], "4.22 Quality Objectives" [p.15], etc.)

Lesson 4.2e: Whenever a highly formalised or formulaic methodology is applied in analyses (particularly in relation to quantitative data), great care must be taken that the methodology is used validly, and does not start to subtly constrain what conclusions may be reached and the transparency of assertions made.

From the perspective of a reader, the provision of obtuse, jargonistic or overly complex analyses is closely akin to just supplying the raw data in an annex, in that the reader feels like they are being 'led' to a conclusion, rather than being *convinced* of it.

Evaluators with a particular area of expertise must also ensure that they do not over-apply analyses drawn from that field at the expense of appropriate (and perhaps much more basic), consideration of other relevant factors[18].

In addition, the more complex and formulaic methodologies become, the easier it is to make apparently minor, but far reaching errors in their application. It is a very short step from non-rigorous application of supposedly mathematical analyses to serious logical errors. The great danger being that the figures produced as an end-point of complex methodologies do not have obvious implications, and this often prevents recognition of spurious results (even by the evaluator conducting the analyses).

Around a quarter of reports in the sample displayed such methodological concerns.

4.2e Example A
From: Evaluation of Vocational Training Projects in Vietnam (December 2003)

The enumerated rating scales used to differentiate between the "Degree of influence of stakeholder on project success" in Tables 1.7 (pp.32/33) and 2.14 (pp.97/98) of this report are not necessarily ordinal in nature. They may therefore be subject to inconsistent interpretation, negating the comparability of 'scores' provided by respondents. The scale is given as:

> **"0 = Without consent**
> **1 = None/little influence**
> **2 = Some influence**
> **3 = Average influence**
> **4 = High influence**
> **5 = Special influence" (pp.33/98)**

Firstly, the initial item is 'extra-dimensional' in that it is not really part of any progressive scale of 'influence'. Secondly, when faced only with the 'narrative explanations' of each of the above scores, it is very possible that some respondents may subjectively order them differently.

---

[18] While it is not suggested that all evaluations must incorporate a full range of appropriate expertise, it is suggested that evaluators are not at liberty to ignore important issues simply because such issues fall outside their normal 'academic comfort zones'.

It is acknowledged that this is unlikely to have been a serious problem in this case, given that the evaluation team did the scoring itself, but the methodological lesson remains valid.

Tables 3.4(a), 3.4(b) and 3.5 (p.110) have more significant problems in that rely on the mathematically invalid practice of 'adding of ordinal scores' to derive an overall 'Mark' (for the desirability of various forms of livestock raising). Regardless of any precedents observed or 'intuitive' interpretation, such approaches are not logically defensible[19]. This part of the assessment must therefore be disregarded and its consideration has been a waste effort by both authors and readers. There is also a danger that, as a general reader may not recognise such technical flaws, the potentially spurious results such invalid practices provide might mislead future decision-making.

On a more general point, while the report does identify omissions relating to project M&E frameworks and introduces the concept of risk management into some analyses (see Section 4.6), it does not critique these omissions, nor explore solutions to any significant degree. For example, in regard to the first project, the report points out:

  "There is no indicator on environmental impact in the project documentation…There has been no study of the impacts on workers' health." (p.23)

   "There is no indictor on water." (p.27).

No recommendations subsequently made specifically address the lack of appropriate environmental indicators[20]. Additionally, for a project that is based on replacing environmentally unfriendly chemical fertilizers with bio-fertilizer, it is disappointing to see that Table 1.6, which incorporates 'Possible Negative Impacts' only addresses potential environmental impacts in so far as to note:

  "The fermented bacterium of bio-fertilizer stinks." (p.27)

Given the technical nature of other analyses in the document, it appears that the likely cause of this shortfall is that neither of the two evaluators had an environmental background. Therefore, there may have been a (perhaps subconscious) favoring of familiar fields (viz. Economics and Sociology), resulting in discounting of other important considerations.

4.2e Example B
From: Community Education Programme Evaluation – Cambodia (July 2005)

This evaluation uses a number of participatory data collection techniques, two of which involve the eliciting of ordinal ratings/scores from respondents. The first, the 'Data Analysis Matrix' avoids the common pitfall of 'adding and/or averaging ordinal ratings' by directly obtaining 'consensus scores' from groups. The second technique, the 'Scale Questionnaire', is not so fortunate and hence the graphs derived from this data, and all findings interpreted from them are not valid results.

Two other techniques used were community group generation of Venn Diagrams (showing social

---

[19] A good reference on this point is the LEARNING TECHNOLOGY DISSEMINATION INITIATIVE (LTDI) *Evaluation Cookbook: http://www.icbl.hw.ac.uk/ltdi/cookbook/info_likert_scale/index.html*

[20] Under the heading of 'General Conclusions' the evaluators do generally suggest: " *Project planning, with a system of well-defined monitoring indicators, is an area for possible improvement in order to ease other steps in project-cycle management, particularly in monitoring and evaluation.*" (p.XV)

[21] Although a breakdown of responses by stakeholder groups is provided and gives a valuable separation of different perspectives, qualification of entries which amount to a particular group commenting on its own performance is absent.

[22] E.g. it may be of little use to ask an illiterate peasant farmer for his/her 'Economic Rate of Return' [example drawn from experience outside sample]

[23] E.g. a peasant farmer who grows crops, uses the by-products to make home handicrafts, and personally takes surpluses of either to sell at local markets, should never be asked whether they fall into the category of 'producer' OR 'manufacturer' OR 'marketer', without the ability to select more than one option [exmple drawn from experience outside sample, excuse given was that: "*if you didn't make the options mutually exclusive, you couldn't easily feed the resulting data into SPSS Statistical Software*"]

networks) and 'Asset-Mapping' (usually by more select groups). Both of these valid qualitative techniques recorded important issues in associated discussion notes.

Of the four techniques, only that associated with the 'Scale Questionnaire' is inherently invalid and no objection whatsoever is made to the inclusion of the others. However, application of the techniques used requires additional comment.

Firstly, interpretation of graphs derived from the ratings techniques is somewhat mechanical, identifying only overall trends, and often failing to discuss potentially interesting (but obvious) exceptions. Secondly, the results lack necessary qualification. This is particularly important in relation to rating analyses from the Data Analysis Matrix, because although responses to such scales are inherently subjective, this is not always obvious in the presentation of data[21]. Thirdly, and most importantly, the results of the four different techniques are not contrasted or discussed as a whole, despite some obvious inconsistencies between them. For example, graphs consistently show highly positive measures of "Increased Local Authority Support" (as perceived by the community), but discussion notes from the more qualitative techniques included partially contradictory issues, such as:

   "Threats from Police." (p.6)

   "Satisfactory relationship with commune, but not with Police, because on some occasions, the local authority may be a relative of the perpetrator." (p.12)

As such comments are received from two of the three locations considered, this should have indicated that some 'unpacking' of the general finding is warranted. This is not provided in the form of an overall assessment of the results of all analyses, nor is it considered in the 'Conclusion' section. Using a range of overlapping analyses is of little value if results are not compared or contrasted at some stage.

4.2e Example C
From: Evaluation of Early Childhood Care and Development Project – India' (May 2005)

Survey design should never be taken lightly. Simply compiling a set of 'reasonable sounding' questions to target 'what the evaluation wishes to find out' is often a recipe for obtaining spurious data.   Many subtle factors must all be carefully considered, such as:
- the reasonable ability of chosen respondents to 'understand[22], know or recall';
- the need for response categories to reflect reality, not constrain it[23;]
- the likely biases in responses of groups (directly or indirectly) evaluating themselves; and
- the possible need for confidentially (e.g. having a field on the top of the questionnaire along the lines of "Respondent: (Put your name here)" [p.21], may immediately constrain the frankness of information provided).

However, possibly the most fundamental caveat is not to ask 'leading' questions. For example:

   "How has this learning (in your view) improved PU's [Partner Programme Units] capacity to program, manage, upscale and review their ECCD [Early Childhood Care and Development] programs?" (p.21)

   "How have Technical Support Organisations (TSO) contributed to improved ECCD programming (including its acceptance) at various levels of the organisation (including the PUs)?" (p.21)

Clearly, some researcher assumptions and convenient respondent prompts have crept into these questions.

Lesson 4.2f: Overall methodologies must be responsive to preliminary findings and allow for an adaptive focus of component analyses.

Evaluators should not be inflexible in regard what key issues an evaluation considers. They must not let their expectations predetermine the total range of analyses deemed

'necessary' under a particular evaluative methodology. Preliminary investigations will often bring surprises requiring the reprioritisation of planned considerations. Unforeseen critical factors may render some intended analyses completely redundant and demand incorporation of new ones. While larger/longer evaluation exercises clearly have more scope in this regard, even small studies must remain somewhat adaptive, if they are to avoid potential irrelevance.

The real frequency of occurrence of this problem is impossible to assess, as it is only when incidental warning signs are present that it can be diagnosed. The fact that such warning signs were found in over 10% of reports is cause for concern.

4.2f Example A
From: Evaluation of Vocational Training Projects in Vietnam (December 2003)

The following observation is made in regard to the first of the three projects under consideration:

 "The number of users as well as the quantity of bio-fertilizer used peaked in 2000 and dropped dramatically in 2001, or right after the project. This could partly or wholly be explained by the close of the factory*, or the supply, while there were no other alternatives in the local area." (p.VII) (*project supported)

There is some very cursory later discussion on this issue:

"There were two reasons to close the factory down. At the end of the year 2001, the price of chemical fertilizers was lower than that of bio-fertilizer, meanwhile the district women's union was unable to buy peat (the main material for production)." (p.13)

However, as it amounts to a key sustainability block, it might have reasonably been raised from the status of a passing comment to a main line of investigation. All the other (highly technical), economic and social analyses provided as to the potential benefits of <u>use</u> of bio-fertilizer are inconsequential, if supply is not assured.

It is also interesting that parallel items also appear in the sections of the main body of the report referring to the second and third projects under examination, with even less attention paid to them in analyses:

  Second project:

"After training, 100% of participants started production, but due to lack of some raw materials some households stopped." … "Lack of raw materials." …"Difficulty in materials."…"Due to difficulties in raw materials, some households had to purchase from outside." (all Table 2.3, p.55)

  Third Project:

"Lack of Materials" (three such entries in Table 3.10, p.119)

"Nearly 80% of the communes can not maintain the number of producers, mainly due to a lack of raw material." (p.119)

It would appear that 'supply of raw materials' was, in fact, a common issue of importance across all three projects of the joint evaluation. The issue might therefore justifiably receive a great deal more focus, not just in the individual project analyses, but also in an overall context. A valuable opportunity to raise the profile of this critical issue and examine it more thoroughly is thus lost.

Lesson 4.2g: The existence of contradictory statements or data which are ignored in discussions, or inclusion of incidental references which themselves suggest important issues that are never raised, are a sign of an incomplete evaluative process.

The most obvious case of this occurs when a graph or figure provides a mixture of very different trends, but only the most broad generalisation is presented of the data (see also 4.2e Example B). While broad generalisations often need to be made, they should not go unqualified, especially when the reader is faced with information suggesting many unexplored exceptions[24].

Limited or incidental references to matters which some readers may regard as of great importance are also of concern. If an item is deemed to be worth mentioning, some explanation of it is necessary for a reader to place it in appropriate perspective.

While this is another pitfall which may be impossible to avoid entirely, around half of the sample displayed the practice beyond a 'reasonable' limit.

4.2g Example A
From: Evaluation of Vocational Training Projects in Vietnam (December 2003)

The following is provided as an isolated comment, but raises significant issues as to the effectiveness and motivation of those involved in extension work:

  "It should also be noted that most of those who have propagandized have not used bio-fertilizer because they have not trusted the quality and effectiveness of bio-fertilizer." (p.14)

A more detailed examination of the issue is clearly warranted.

4.2g Example B
From: Indonesia Papua Health and Education Programme (March 2005)

While there is no one reference in this report that creates concern, the general reporting of failures in regard to local control and monitoring of financial aspects of the project leave the reader speculating as to whether any 'impropriety' is suspected by the evaluators. Such concerns are of great significance to implementers and donors alike, and should never be left hanging. While it is inappropriate (and potentially defamatory) for evaluators to make pointed accusations in this regard, there are a couple of 'standard fallback options' available to authors who find themselves in this situation: they may either clearly justify why 'impropriety is <u>not</u> suspected, or alternately, may recommend that 'an immediate financial audit be made of project funds'.

4.3 The drawing out of 'lessons learnt' and their usefulness

Identifying 'lessons' is a key element of the continuous improvement role of evaluations. While such lessons are often implicitly available to a reader from the outcomes of analyses, taking the additional step of explicitly formulating and presenting them in reports has obvious benefits.

---

[24] Bearing in mind it is often such exceptions which generate the most useful lessons learnt, and in certain cases, these references may constitute the warning signs for 4.2f.

However, there are a few caveats on this generalisation. 'Lessons' need to be both insightful and useful. As with recommendations[25], lessons which are extremely obvious or 'motherhood' in nature risk both insulting the intelligence of the reader and providing little practical guidance on how they should be applied across varying contexts.

To expand on the latter point, evaluations are context-specific studies and their authors are often not in a sufficiently informed position to generalise findings. It is therefore appropriate that they largely limit themselves to extracting lessons which are well-grounded in the details of situations examined[26]. While this potentially limits the breadth of application of such lessons, it also limits the breadth of potential **misapplication** andmaximises usefulness to those working in the field.

Just under half of the sample either didn't explicitly extract lessons, or provided (some) lessons which were difficult to apply.

4.3 Example A
From: Evaluation of Vocational Training Projects in Vietnam (December 03)

While some lessons learnt 'headings' are included in the main body of this report, what follows are often not clearly 'lessons' (i.e. items prompting improved performance in future), but rather are basic findings. For example:

   "The right use of bio-fertilizer can make high economic impacts, increase income and decrease production costs" (p.19)

4.3 Example B
From China Integrated Development Programme Evaluation Report (March 2005)

The following are examples of general lessons provided in these reports:

   "Lesson: Exploration and analysis is required to learn lessons as to how to advance from pilot projects." (p.4)

   "Lesson: Good management systems and professional expertise ensure good implementation." (p.4)

Note that all of these lessons are significantly removed from the context of the project being examined. The second example also illustrates the danger of extrapolating from 'the specific to the general', in that, while 'good management systems and professional expertise' are possibly necessary conditions for satisfactory project implementation, they are often far from sufficient conditions.

4.3 Example C
From: Ex-Post Evaluation Report: The Savannakhet Disaster Preparedness Project - Lao PDR (April 2005)

This evaluation explicitly provides 'lessons learned', but does not include explicit recommendations. This may be largely justified, given the fact that this is a truly 'ex-post' study, with no immediate project stakeholders remaining to whom recommendations might be made[27]. Some of these lessons provide good examples of provision of an appropriate level of contextual detail to ensure usefulness, even without a highly specific target audience in mind. For Example:

---

[25] In fact, 'lessons' might well be defined simply as recommendations which target a wider audience (i.e. beyond the stakeholders of the project in question).

[26] In short, the lessons derived should be lessons for similar projects.

[27] Although it is arguable that some of the 'lessons' which more specifically targeted the implementing agency might have been more appropriately called 'recommendations'.

> **"It may be difficult to adequately capture the desire for wells, either because non-productive inputs may seem outside the scope of an agricultural/disaster project and/or because women, who are the major constituency for wells, are usually less able to get their priorities represented in village-wide planning." (p.21)**
>
> **Contrast this against the lessons from the preceding example. Note that this lesson goes well beyond a motherhood statement or 'noble desire'. It is sufficiently contextual to alert workers in similar situations to its relevance.**

4.4 Clarity and conciseness in presentation

It is not just analyses themselves that need to have a clear and logical structure, the reports in which they are imbedded should also reflect such rigour. Key factors include the excessive use of acronyms or specialist jargon[28]. While authors may mistakenly believe this makes them look well informed or 'technically up-to-date', from the readers' perspective it invariably appears inconsiderate. Any document that is not easy to read and understand displays the most basic failure of authorship.   While no particular standardised structure is suggested here, a few rules of thumb should be considered. These include:

- Adopt a reasonable (i.e. task-sensitive), 'indicative page range[29]' which ensures both adequacy of reporting and prioritisation of content. Authors need to recognise that most readers have limited time available for digesting reports, so 'comprehensiveness' will often need to be traded off against ensuring presentation of priority issues[30];
- Provide an initial summary of findings (relating to both accountability and continuous-improvement issues) for documents over 10 pages[31] (6 of the 19 relevant reports in the sample did not do so);
- Ensure that the document structure is in sympathy with the structure and relative importance of underlying analyses used; and
- On completion of a draft report, globally replace as many acronyms as space permits, and provide a summarised list explaining those that remain (7 of the sample reports did not do so).

Just under half of the sample display clarity and conciseness concerns of one form or another.

> **4.4 Example A**
> **From: Sustainability Evaluation Takeo/Kandal Project – Cambodia (March 2005)**
>
> **This report does not explicitly differentiate recommendations from less action-oriented findings. While statements in the form of recommendations are made (and are generally well supported), they are not presented in a manner which promotes their easy identification. For example:**

---

[28] See also 4.2e.

[29] Note, the concept of a 'page limit' is valuable, provided a minimum size is also made explicit.

[30] This prioritisation is a key 'professional' task of the evaluator.

[31] These are often termed an 'executive summary', but need not take a full (or multiple) page format in the case of smaller documents.

"(39) The key issue is whether or not the VGG [Village Graduation Guideline] tool adequately addresses issues of ongoing sustainability, particularly of the VDC [Village Development Committee]. It is the view of the study team that it does not, and therefore either the current tool needs to be strengthened in this area or a 'village graduated sustainability tool' be developed. For reasons of consistency the team believes the best option would be to enhance the existing tool." (p.13)

"(45) On the balance the study team believes that the intrinsic weaknesses in the broader context in which the Takeo/Kandal project has operated requires further involvement by LWF [Lutheran World Federation] beyond 2005." (p.14)

These recommendations are imbedded within a large pool of general discussion paragraphs (similarly numbered) from which they are not differentiated either by labeling or structure.

4.4 Example B
From: Evaluation of Mongolian Red Cross Society Strengthening Public Health (HIV/AIDS and Blood Safety) (July 2005)

In the main report over 40 recommendations are made, the executive summary only presents 11 of these and makes no explicit reference to the existence of others. While it is appropriate that an executive summary be a concise overview of the findings, it is dangerous to attempt to summarise recommendations themselves. Even when a selective list of what are seen as 'Key Recommendations' are presented, care should be taken to explain that this is an abridged version of the full set of recommendations made.

4.4 Example C
From: Evaluation of Vocational Training Projects in Vietnam (December 2003)

The presentation, and to a significant extent content, of recommendation sections within the main body of this document (pp.38/39, p.85, pp.127/128 and pp.142/143) are not consistent with that provided in the corresponding sections of the executive summary (p.IX, p.XII and p.XIV). The executive summary versions are not consolidations of the former, but often contain more detailed, more explicitly stated, and even new recommendations. This is confusing from a reader's perspective.

Further, this report is over 140 pages in length. While it contains some robust analyses, it also contains some flawed approaches[32]. However, its main point of concern is that these analyses are not always focused on key topics, and it does not marry the results of (often divergent) analyses into a consistent set of well-argued findings. Imposition of a reasonable page range, and the subsequent prioritisation of analyses included in the report would clearly have benefited both authors and readers. For a converse example (i.e. a report which is 'too concise'), see 4.2a Example A.

4.5 Independence of evaluators

While one might expect this point to refer to whether evaluators are drawn from within the implementing agency or externally, in the case of the sample reviewed[33] this does not appear to be a key determinate of the perceived 'independence' of findings. There are cases of both internal and external evaluators suffering from a tendency to focus on the most favourable findings, but equally, there are many well balanced reports across both author types. Whether there has been some degree of 'client capture' of external evaluators is beyond the scope of this metaevaluation to determine. However, the fact that many internal reviewers were balanced in their assessments is undeniably praiseworthy. Perhaps a more important consideration in relation to this sample is the

---

[32] See 4.2e Example A.
[33] Or at least for the subset that provided the relevant information on evaluation team membership.

issue of **conflict of interest**. This involves the evaluators taking on roles which are inconsistent with remaining 'impartial'.

While only a couple of reports show clear conflicts of interest, it should be noted that more than a quarter tend to couch even the problems encountered in the most positive light feasible.

4.5 Example A
From: Evaluation of Mongolian Red Cross Society Strengthening Public Health (HIV/AIDS and Blood Safety) (July 2005)

A paragraph from the report contains the following excerpts:

  "Concerns were raised by the National Aids Foundation (NAF), the International AIDS Alliance partner in Mongolia, in relation to perceived overlap with their projects supporting MSM (Men who have Sex with Men) and Positive people." … "We were also able to clarify with NAF's Director that MRCS was working in line with global Red Cross HIV/AIDS strategic directions and would make every effort to respond to requests for support and to engage in partnership with positive people and vulnerable groups, while acknowledging the need to coordinate efforts and keep each other informed of developments." (p.8)

Given the wholly internal composition of the evaluation team, the response provided is 'understandable' and may seem like a minor piece of opportunism. However, any evaluation exercise must maintain both a sincere and perceptional standing of impartiality. Any attempt by the evaluation team to 'defend' or 'justify' the project or any of its key stakeholders (especially implementers), clearly threatens such standing, and may be regarded as evidence of conflict of interest. (Note: while it is appropriate for the implementer to eventually try to allay fears raised by respondents to the evaluation, it is not appropriate to use the evaluation process itself to do so).

4.5 Example B
From: Faith Mission Health and Education Services Review (July 2005)

Firstly, in stating the purposes of the review, the report contains the following excerpt:

  "Firstly, it attempts to reflect the impact of the activities carried out by FMHE [Faith Mission Health and Education Services] along with all the lessons learnt and then use the information to develop a concept paper for FMHE to use as a marketing tool to gain support for its work." (p.7)

A review of the formal TOR for this evaluation does not appear to require development of such a 'marketing tool'. Regardless of where this purpose originated, it is clearly not consistent with the conducting of an impartial assessment of the project under scrutiny.

Secondly, in a section regarding development of an exit strategy[34], the suggestion is made to undertake an annual "Joint Organisational Assessment". One of the risks of doing so is given as:

  '-reports may be used against the organisation.' (p.33)

Other than from a workload sense, transparency measures need to be regarded as project-neutral. Their value should then assessed in relation to how they form part of an effective M&E framework; which will rightly reveal both positive and negative aspects of the project. While a 'risk' in an absolute sense, the above example is one which it would be improper to consider. (N.b. Despite these inclusions, no bias is actually shown in the main content of this report).

---

[34] Commendably, it was one of the few reports to include this.

4.6 Level to which project Monitoring and Evaluation (M&E) frameworks are considered, particularly in regard to Risk Management

Evaluations form one element of a project's M&E framework. An evaluation report which makes no reference to this greater whole (especially in relation to the value of the information gathered from more routine monitoring exercises), not only misses an opportunity for more informed judgment, but neglects important risk management aspects of project implementation.

While legitimately supplemented with more ad-hoc or opportunist information gathering, a robust project M&E system should include indicators of the status of predictable risks. This does not just mean an ability to recognize poor showings on direct indicators of success; specific indicators should also be developed which reveal whether the key assumptions upon which success is reliant continue to hold[35].

Just under half of the reports fail to significantly address M&E issues, and only two formally address associated risk assessment (though neither to an exemplar level).

4.6 Example A
From: Kariobangi HIV Project Evaluation, Kenya (June 2004)

This report provides a very positive example of appropriate consideration of the position of a particular evaluation in regard to the overall M&E framework of the project. It includes a table (Table 1.2 p.8/9) listing the recommendations of previous reviews, and noting the progress made against them. The Participatory Evaluation of the Indonesian Red Cross HIV Hotline Counseling Service (June 2005) is the only other report in the sample to contain a similar section.

4.6 Example B
From: Ex-Post Evaluation Report: The Savannakhet Disaster Preparedness Project -Lao PDR (April 2005)

One of the 'lessons' of this report is worth quoting, not just because of its sufficient attention to detail, but because it also demonstrates that the 'critical' perspective of an M&E exercise (such as an ex-post evaluation), can legitimately extend to the findings of former evaluative exercises. The greater 'lesson' being that evaluators should take no bases or opinions 'on trust':

  "The final evaluation did not comment on the limitations of the indicators, nor on the failure to carry out DVA [Disaster and Vulnerability Analysis] as planned. Furthermore, the final evaluations that were carried out avoided directly confronting these limitations and omissions in the project indicators and M&E, thereby missing an opportunity to improve the capacity of CARE Laos in project design, as evaluations should do. So the lesson is that evaluations should rigorously apply the LogFrame and draw attention to problems, and CARE Laos should see all evaluations as capacity building exercises" (p.21)

4.7 Attempts to move beyond output-level analysis, to also attempt to consider modest 'outcomes'

It is emphasised that in the *project* evaluation context, 'outcomes' are not regarded as particularly high-level considerations. They are simply those impacts of delivering

---

[35] Or, put conversely, whether potential threats to success have eventuated.

outputs which may be reasonably extrapolated or measured[36] from within the reference frame and data collection capacities of the project. Hence, they are unlikely to reach the 'Goal', or even the 'Purpose' levels provided in a project logframe[37]. This 'intermediate' nature of such outcomes does not reduce their potential usefulness in understanding project success or failure, so despite this qualification they remain an important topic for consideration of ANCP evaluations.

Just over half of the sample failed to address even modest outcomes to a significant degree.

4.7 Example A
From: Kariobangi HIV Project Evaluation, Kenya (June 2004)

This report contains an excellent example of an analysis which tries to look beyond simple outputs delivered and report on 'outcomes'. It also illustrates the level of effort this requires, if robust findings are desired:

In this evaluation, a sample of 1250 households is subjected to an exhaustive set of information gathering techniques, including modest medical assessments. These 1250 households include 352 that are not involved in the project, so that a comparative 'control' group could be approximated. The approach produces large amounts of data, both quantitative and qualitative. Both the directly quantitative data and the frequencies with which qualitative responses are returned are fed into statistical analyses, such that the statistical significance of results could be reported[38].

While few evaluation exercises could afford the resources to match the scale of this study, it remains an example of the rigour and objectivity required to produce verifiable results; as opposed to results which equate to 'considered opinions' of evaluators. Those commissioning reports with TOR that include a requirement to assess 'outcomes' or 'impacts' must therefore provide commensurable resourcing. Otherwise they should temper their expectations in terms of both how far above output-level such outcomes may be considered, as well as how verifiably objective will be the extrapolations made to reach them.

4.8 Use of participatory techniques

A number of the evaluations in the sample use participatory techniques[39] as a means of collecting information. As with approaches to evaluations themselves, the available range of valid participatory techniques is huge and it is clearly inappropriate for this metaevaluation to favour specific ones.

However, caution must always be exercised in using the subjective perspectives obtained from participatory techniques. Such techniques should never be confused or equated with the subsequent *analysis* of the data they provide[40]. One concern arising from the sample was that, in some instances, information gained in a participatory manner was given 'special status', in that it seemed to be exempt from the rigor of 'critical review'

---

[36] With a satisfactory degree of attribution.

[37] Which usually require evidence collected by dedicated research at programme-level.

[38] Although it should be noted that (as with **4.4 Example C),** despite the inclusion of such a rigorous analysis, this document largely failed to bring all the information gained to bear on producing and justifying a coherent set of recommendations and conclusions.

[39] Defined for the purposes of this metaevaluation as any means of encouraging stakeholder involvement in the evaluation, beyond a basic standard of one-on-one interviews or simple survey forms.

[40] The participatory technique should most often be regarded as a *data-collection* exercise.

routinely applied to other sources. An evaluator must always be aware that all societies and cultures contain just as many instances of self/vested-interest, ill-informed belief and desire to please as their own. To the extent that political sensitivities of report circulation allow, analyses of the effects of such factors should always be included as qualification to apparent results.

Of those that substantially use participatory techniques, virtually all make the error of mistakenly equating participatory data collection techniques to a complete 'analysis'.

4.8 Example A
From Kariobangi HIV Project Evaluation, Kenya (June 2004)

Despite including one of the most robust and 'objective' analyses of the entire sample, this report fails to maintain this level of rigor in developing its recommendations and conclusions. The recommendation section begins with:

  "In a participatory context, the recommendations also include those that were obtained from the respondents." (p.47)

While some of these participant recommendations are separated into a table (Table 4.1 p.48), the section then continues:

  "Other recommendations that accrued from the qualitative data are also listed…" (p.48)

Many of the remaining recommendations do then appear to be drawn from many different sources and contexts, rather than from a single consistent analysis. Directly presenting an unfiltered/unqualified list of 'good ideas' solicited from stakeholders amounts to a lapse in professionalism in providing analysis. (Note: It is not suggested that participant input shouldn't be included in analyses, just that it needs to be appropriately qualified with objective criticism or support). This example was not an isolated one within the sample.

---

5. USE OF EVALUATION ALLOWANCE BY ANCP NGOS

(N.B. Considers all ANCP NGOs, not just those contributing sample evaluations)

5.1 Data Issues

As a brief desk study, the metaevaluation relied on information contained in files for the financial years 2004/5-2006/7. This primarily consisted of ADPlans or ADPlan Reports. At the outset of the study, it was decided to calculate expenditure percentages based on original Indicative Planning Figure (IPF) allocations (rather than end of year figures) in recognition of the fact that evaluations may require significant planning and lead-in times.

Despite this simplified approach, the calculation of percentage expenditure on evaluation proved extremely problematic due to inconsistencies in how usage of these funds was reported. Many older ADPlans have no separate evaluation line item in budgets, or use combined line items (e.g. 'Monitoring & Evaluation', 'Monitoring, Evaluation & Training' or 'Administration and Evaluation')[41]. Even in later ADPlans, while there are evaluation line items included in budgets, there are often no subsequent reports referenced elsewhere. There are also a number of occasions of no evaluation budget line items (or other references) in ADPlans, but copies of completed evaluations on file. In short, on the basis of the available information, it is risky to suggest which organisations are using (or not using), the ANCP option (up to 5% of total IPF) for funding evaluative activities.

5.2 Tentative Quantitative Findings

Some agencies appear not to have undertaken formal evaluative activities using ANCP funds in the financial years 2004/5-2006/7. The available data (see section 5.1) suggests that this amounts to approximately 30% of those receiving ANCP funds. Of those utilizing the option, about 60% allocated less than an average of 5% per year, and under 10% exceeded the 5% per year average.

Clearly all agencies should be doing some evaluative assessment of their current or past work[42]. It is difficult to be critical of those NGOs which are already utilising the ANCP funding option, whether above or below the limit, because multi-year project cycles could rightly influence the levels of such expenditures.

5.3 Qualitative Findings

In contrast to the quantitative findings, significant qualitative inferences could be reliably made from the available information. Many of these pointed to a need for a clearer understanding of what constitutes an 'evaluative exercise' upon which the ANCP 5% option may be legitimately expended. Rather than try to develop a highly prescriptive

---

[41] The latest ADPlans (06-07) require a separate line item.
[42] This requirement is inherent in the accreditation process.

definition, the metaevaluation proposes a pragmatic set of *principles* which might be adhered to. These are illustrated using the examples below:

5.3.1 What may be classed as 'Evaluation' for the purposes of ANCP?

In certain ADPlans which included a budget line item for 'Evaluation', the level of annual expenditure recorded creates some concern. Some such inclusions are well under $500 (the record being $40). Given that any reasonable evaluative exercise should include some defensible level of dedicated investigative effort, these very low expenditures suggest a vague interpretation of what may be charged against an 'evaluation' line item.

It is recognised that the term 'evaluation' may be validly used as either a noun (i.e. an 'evaluation') or a verb (i.e. the process of 'evaluation'). It is therefore not the intent of the metaevaluation to impose an overly constrained definition. Rather, a few practical rules of thumb are again suggested:

- An evaluative exercise must consider the project (or program) as a whole, not selected aspects of it[43]; **AND**
- An evaluative exercise must consider both accountability and continuous improvement issues; **AND**
- An evaluative exercise must include both data gathering/compilation[44] and analysis. This may target qualitative or quantitative data, but should not amount to an unqualified presentation of subjective opinion; **AND**
- An evaluative exercise must produce a written report of some form; which includes conclusions and recommendations[45] .

Despite the above stipulations, evaluative exercises may still include a wide range of activities. Mid-term or other periodic *reviews* of various sizes remain just as valid as the 'typical' (i.e. large scale, ex-post, external) interpretation of an 'evaluation'.

Note that the ANCP program now makes a distinction between 'monitoring' and 'evaluation'[46]. Monitoring tasks are no longer to be funded under evaluation allowances, nor are combined (M&E) budget lines now permitted. To make the distinction between the evaluation and monitoring more explicit (without a lengthy discussion of the subtleties necessarily involved), it is suggested that any exercise which meets (ALL) the above rules of thumb[47] may be validly regarded as 'evaluation'. Information gathering

---

[43] Some ADPlans include 'evaluations of training exercises'. Such selective assessments may not inform (and may indeed misinform) appropriate future action in relation to the project as a whole (e.g. you may be training people very well, just not in the skills the project needs them to learn!).

[44] Note data from regular monitoring may be used, but this should be independently augmented, where possible.

[45] Some ADPlans include references to 'evaluative workshops'. This is acceptable provided an associated report with recommendations covering the full scope of the project are an output of such workshops.

[46] As of 2006.

[47] And in so doing appropriately addresses the issues raised in Section 4.

alone, or the partial/incidental analyses of such information should be considered 'monitoring'.

5.3.2 Balance of Summative vs. Formative Evaluative Exercises

Despite the general trend, a number of organisations do report very exhaustively on their expenditure of ANCP funds for evaluative activities; to the extent that they have developed separate ADPlans solely for their annual 'Evaluation Programs'. While there is nothing necessarily wrong with this approach, it was noted that such evaluation programs tended to consist only of large (often single), summative[48] evaluations.

Given the limited amount of funds available each year from the 5% allowance, it does make apparent sense for ANCP NGOs to consolidate evaluation expenditure across all their projects in this way, to ensure that *some* more exhaustive studies take place. However, this approach should **not be followed to the exclusion of perhaps smaller, but more regular, formative[49] evaluative exercises** (e.g. mid- term reviews).

The DAC has also identified this tendency within the wider development community as a 'too much, too late' syndrome in recognition of a number of factors, including that focusing on post-completion assessments alone may mean that fewer opportunities are created to 'fix' or improve projects as they are implemented[50].

In addition, removing evaluation components from the ADPlans for specific projects tends to inadvertently suggest that evaluative processes are an 'added extra', rather than an integral part of the project's overall M&E framework. Given the 'typical' interpretation of 'an evaluation', this again tends to distort the balance between summative and formative assessments.

The principle suggested by this metaevaluation is simply that a reasonable balance needs to be maintained between formative and summative evaluative exercises. In practice, this is likely to mean regular (perhaps smaller), formative exercises augmented by occasional (perhaps larger), summative exercises.

5.3.3 Efficient use of available resources

There is no denying that many of the projects undertaken using ANCP funding are attempting to meet the most basic needs of beneficiaries. It is understandably difficult therefore, for some to justify allocating scarce resources to non-operational purposes such as evaluation. Hence, it is appropriate that the funding option under the ANCP program allows up to 5% of ANCP subsidy to be used, rather than presenting a fixed target. This is especially relevant to base-funded NGOs for whom the total amount of subsidy may be significantly less than $100 000.

---

[48] I.e. feeding into next project cycle: end of project or ex-post studies

[49] I.e. feeding into current project cycle.

[50] Taken to it's logical conclusion, this might mean that an agency eventually becomes '*very good at telling how badly it has been doing*'. There are also concerns that by the time an ex-post evaluation is completed, the implementation context may have substantially changed, rendering lessons out-of-date.

However, it is unwise to simply go on completing 'tasks', without occasionally reflecting on the 'whole picture'. As most development professionals will know, components of projects naturally tend to 'take on a life of their own', and without appropriate care the synergies of integrated project design can easily be lost. It is therefore important that NGOs consider the 5% allowance when initially developing their ADPlans, such that it is added to, rather than extracted from, operational costs. Unfortunately, the great majority of base-funded agencies are already receiving their full $100 000 limit and are not in a position to do this.

One option for funding occasional major evaluations by NGOs with low levels of ANCP subsidy would be to set aside a separate 'major evaluation fund' which is not allocated through a particular ADPlan (perhaps $50,000: allowing two major evaluations per year). AusAID could then ask NGOs below a certain funding threshold (perhaps $200,000), to indicate in their ADPlans which project they would choose to evaluate if funds were made available. The choice of the major evaluations funded each year would need to be made at the time ADPlans are approved, and could be based on an organisation or sectoral 'roster', designed to reflect the learning priorities of AusAID, as well as ANCP NGOs. This approach is suggested as it is relatively low cost, and would not mean major change to existing systems.

It is stressed that this approach needs to be considered in addition to the current 5% allowance. It would effectively provide an alternate option for funding occasional major reviews/evaluations and hopefully then allow the existing allowance to be more effectively used to restore a balance including smaller, but more regular, formative evaluative exercises.

For those NGOs with smaller subsidies, rolling-over a proportion of the evaluation allowance into future years is an option for funding periodic reviews[51] which may already be applied for. Making the process of rolling over unspent portions of the allowance easier would help encourage such practices.

5.3.4 What to Evaluate?

The information provided in ADPlans also reveal some interesting choices made in regard to which projects are subject to evaluation. Some agencies, faced with more than one project under ADPlans, and a very limited amount of money to undertake an evaluative exercise, simply chose their smallest option. Presumably, it was felt that the smallest project was the only one that could be reasonably assessed for the available funds.

While the apparent logic of such decisions is recognised, there is some false economy involved. For an organisation to realize the greatest benefit out of any learning exercise it is crucial that it give priority to evaluating its *core business*. In the vast majority of cases, core business is reflected in an organisation's largest projects.

---

[51] It should be noted that those wishing to do this must explicitly allocate such funds in ADPlans.

6. CONCLUSIONS AND RECOMMENDATIONS

6.1 Evaluation Objective A (Determine the quality of a cross-section of ANCP activity evaluations) & Evaluation Objective B (Determine the validity of a cross-section of ANCP activity evaluations' results and recommendations)

These objectives were considered jointly by the metaevaluation. In regard to the eight key quality issues identified in the sample evaluations, the following (rough) proportions of reports displayed significant **shortcomings**:

1. Level to which the evaluation provides a balance of *Accountability* and *Continuous Improvement* Considerations – **one quarter**;
2. Level to which assertions made (especially final conclusions and recommendations), are supported in a defensible and methodologically sound manner – **just under half**;
3. The drawing out of 'lessons learnt' and their usefulness - **just under half**;
4. Clarity and conciseness in presentation - **just under half**;
5. Independence of evaluators – **one quarter**;
6. Level to which project M&E issues are raised, particularly in regard to *Risk Management* - **just under half for M&E (only one tenth addressed Risk)**;
7. Attempting to move beyond 'output-level' analysis, to consider modest 'outcomes' – **just over half**; and
8. Use of participatory techniques **– Almost all of the evaluations that substantially used participatory techniques displayed shortcomings in their analysis.**

Most of these shortcomings are very basic, and therefore easily addressed by making future evaluators aware of the potential pitfalls related to each issue. The sample evaluations provided valuable illustrative examples.

Recommendation 1. Organisations receiving ANCP subsidies be required to provide a copy of this report to anyone tasked with organising or carrying out future evaluative exercises; including both internal staff and external contractors.

Recommendation 2. ACFID (Australian Council For International Development) be asked to provide copy of this report to anyone commissioned by them to provide collective M&E training to the NGO community, so that it may be used to help prioritise content of such training.

Recommendation 3. AusAID include in the TOR of the next ANCP metaevaluation a requirement to consider whether improvement has been made in regard the 8 key quality issues identified in this report.

6.2 Evaluation Objective C (Determine how broadly and consistently evaluation funds are being utilized by accredited NGOs)

This objective was attempted in relation to the financial years 2004/5-2006/7) of ANCP subsidies, but proved difficult to reliably fulfill given inconsistencies in reporting. In

broad terms, it appeared from the available data that approximately 30% of ANCP NGOs were not undertaking evaluative activities of any kind. Constraints were identified in regard to the shared understanding of 'evaluation' and hence, what may be funded using the ANCP 5% allowance. Some ANCP NGOs appeared to display lack of balance between smaller and more regular formative evaluative exercises and 'occasional' major summative evaluations, with a bias toward the latter. It was noted that the ANCP 5% allowance alone was not sufficient to encourage NGOs receiving a low level of subsidy to maintain an appropriate evaluation regime, and that associated compromises made were often counter productive.

> **Recommendation 4.** AusAID clarify what it means by 'evaluation', particularly in regard to the distinction between 'monitoring' and 'evaluation' (consideration may be given to using the specific suggestions made in section 5.3.1).

> **Recommendation 5.** AusAID require that the names of all reports produced through evaluation expenditure be provided in 'Significant Outputs' sections of ADPlan Reports.

> **Recommendation 6.** AusAID consider simplifying the process for rolling over unused evaluation funds to future years (including options for NGOs receiving greater than $200 000 p.a.), and establishing a modest ($50,000-$100,000) 'major evaluation fund' (external to ADPlans), from which NGOs receiving a low level of ANCP subsidy can occasionally draw to conduct major evaluative exercises.

Annex 1. Participating ANCP NGOs and Associated Sample Evaluations

1. Anglican Board of Missions (ABM):
  ● **Project/Program Evaluation, Montanosa Potable Water Supply Project/Cordillera Water Program, Philippines (July 2005)**

2. Australian People for Health, Education & Development Abroad (APHEDA):
  ● **Evaluation of Vocational Training Projects in Vietnam (December 03)**

3. Australian Red Cross (ARC):
  ● **A Report on the Impact and Degree of Success that ARC and Vietnam Red Cross (VNRC) can Claim in Relation to the Achievement of the HIV/AIDS Program Objectives (November 2004)**
  ● **Evaluation of the Mongolian Red Cross Society (MRCS) Strengthening Public Health (HIV/AIDS and Blood Safety) (July 05)**
  ● **Participatory Evaluation of the Indonesian Red Cross HIV Hotline Counseling Service (June 05)**

4. CARE Australia:
  ● **Ex-Post Evaluation Report: The Savannakhet Disaster Preparedness Project -Lao PDR (April 2005)**

5. International Women's Development Agency (IWDA):
  ● **Evaluation of Four Projects on the Thai-Burma Border (August 2003)**
  ● **Community Education Programme Evaluation –Cambodia (July 2005)**

6. Lutheran World Federation (LWF):
  ● **Sustainability Evaluation Takeo/Kandal Project –Cambodia (March 05)**

7. PLAN Australia:
  ● **Evaluation of Early Childhood Care and Development Project –India (May 2005)**
  ● **Evaluation of Early Childhood Care and Development Project –Vietnam (April 2005)**

8. Salvation Army Australia Development Office (SAADO):
  ● **China Integrated Development Programme Evaluation Report (March 2005)**
  ● **India Health Appraisal and Rural Empowerment Project Evaluation Report (November 2004)**
  ● **Indonesia Papua Health and Education Programme Evaluation Report (March 2005)**

9. Save the Children Fund Australia:
  ● **Faith Mission Health and Education Services Review (July 2005) {PNG}**

10. Sexual Health and Family Planning Australia (SHFPA):
  ● **Evaluation Report, Women's Action For Change Sexual Minorities Project, Fiji (September 2003)**

11. United Church Overseas Aid (UCOA):
  ● **Rural Village Water Project Evaluation –PNG (December 05)**

12. World Vision Australia (WVA):
  ● **Kariobangi HIV Project Evaluation, Kenya (June 2004)**
  ● **Tamanna Child labour Project Evaluation Report –India (May 02)**
  ● **End of Project Review Report –Phase One, rural Palestinian Women's Vocational and Income Generation Project –Gaza Strip (December 03)**

Annex 2. Basic Design Considerations for Making M&E Easier

It is important to note that development of implementation and monitoring frameworks are so closely aligned that they should be considered indivisible. In short, there is as little point in defining tasks which cannot be effectively monitored as there is collecting information which doesn't really reveal if desired tasks are being achieved. If implementation and monitoring frameworks are not appropriately aligned from the beginning of an initiative, **there is little chance of implementation ever being regarded as successful!** (how will anyone tell?[52]).

To put this another way, two necessary conditions for ensuring that success is 'seen to be' achieved are to clearly define what success will look like (via tailored Performance Indicators [PIs] or other Monitoring Data), and to have in place systems for collecting such relevant information. This sounds trivial, but it is where the vast majority of initiatives fail themselves from their outset. Suggesting 'somewhat applicable' PIs at a later stage is a recipe for failure (and a great deal of costly, excess work).

In practice, this means that those developing plans or proposals shouldn't define an objective without concurrently considering whether it is possible and practical to measure (i.e. Monitor) achievement of it.

Designers also need to be very careful with using macro-level indicators as direct measures of success of any specific plan. This applies to everything from health and agricultural, to economic data. The caution here relates to ample past experience of agencies trying to use macro measures because of their perceived convenience (i.e. they are often readily available), but eventually being badly disadvantaged in reporting of success because of it. A few simple examples are:

- sector-wide agricultural programs that used 'increased national crop production' as a measure of success, but then found themselves in the grip of a drought over the next three seasons…
- Educational health projects that used 'reduced national mortality rates', just before a serious flu epidemic…
- Private sector job creation programs that used 'reduced national unemployment rates', just as the bottom fell out of the world commodity market…

In short, the macro measures generally encompass numerous externalities beyond the scope of any plan to control. In such situations it is very important to have the appropriate micro indicators in place to be able to demonstrate '**how much worse'** the trends in the macro indicator would have been **without** the intervention in question.

---

[52] It is a common misperception that having 'vague' objectives shields managers from scrutiny. All professional evaluators (be they accountants, epidemiologists, or governance specialists) will treat a lack of evidence of success as a default form of failure. Note that this does not necessarily apply to higher level 'goals', as it is often inherently assumed that achievement of stipulated objectives will 'contribute' to such goals.

Annex 3. Terms of Reference for the Metaevaluation

Background

1.   The AusAID NGO Cooperation Program (ANCP) is an annual funding program for Australian NGOs that have been accredited with AusAID. The goal of the ANCP is to subsidise Australian NGO community development activities which directly and tangibly alleviate poverty in developing countries.
2.   Under ANCP guidelines, recognising that evaluation is a normal part of the project management cycle, NGOs can utilise up to 5% of their annually allocated ANCP funding for evaluations.
3.   In 2004 AusAID and the Committee on Development Cooperation (CDC) were presented with a "Review of the AusAID NGO Accreditation Process and Systems". In this review it was recommended that "AusAID gather annual evaluation reports from agencies (funded by the 5% that ANCP allows for M&E) and undertake a meta-evaluation."
4.   With the development of new initiatives under the *White Paper On The Australian Government's Overseas Aid Program* (2006), it is now timely to undertake a limited evaluation of the ANCP funded evaluations and relevant documents to determine their quality and validity. It is also timely for AusAID to examine the broader context around evaluations in order to inform future ANCP directions. In this case this process is termed a "metaevaluation".

Objectives

The objectives of this desk-based metaevaluation of ANCP funded evaluations are to:

A)   Determine the quality of a cross-section of ANCP activity evaluations

B)   Determine the validity of a cross-section of ANCP activity evaluations' results and recommendations

C)   Determine how broadly and consistently evaluation funds are being utilised by accredited NGOs.

Scope and Methodology

The scope of the metaevaluation will include:

a)   Read the following NGO evaluations of their ANCP activities:
*   Anglican Board of Missions; Cordillera Water Program Evaluation (Philippines), Dec 05
*   Australian Lutheran World Service; Planning for increased sustainability (Cambodia), Mar 05
*   Australian People for Health, Education & Development Abroad; Evaluation on Vocational Training Projects in Vietnam, Dec 03
*   Australian Red Cross (ARC); A Report on the impact and the degree of success that ARC and Vietnam Red Cross can claim in relation to the achievement of the HIV/AIDS Program objectives, Nov 04
*   ARC; Mongolian Red Cross Society Strengthening Public Heath (HIV/AIDS & Blood Safety), Jul 05
*   ARC; Participatory Evaluation of Indonesian Red Cross HIV Hotline Counselling Service, Jun 05
*   CARE Australia; Ex-Post Evaluation Report: the Savannakhet Disaster Preparedness Project (Lao PDR), Apr 05

- International Women's Development Agency (IWDA); An Evaluation of Four Projects on the Thai-Burma Border (Burma), Aug 03
- IWDA; Community Education Programme (Cambodia), Jul 05
- PLAN; Early Childhood Care and Development (India), May 05
- PLAN; Early Childhood Care and Development (Vietnam), Apr 05
- Salvation Army Australia Development Office (SAADO); China: Integrated Development Programme Evaluation Report, Mar 05
- SAADO; India: Health Appraisal and Rural Empowerment Project Evaluation Report, Nov 04
- SAADO; Indonesia: Papua Health and Education Programme. Mar 05
- Save the Children Australia; Faith Mission Health and Education Services Review, Jul 05
- Sexual Health & Family Planning Australia; Evaluation Report: Women's Action for Change Sexual Minorities Project (Fiji), Sep 03
- United Church Overseas Aid; Rural Village Water Project Evaluation (PNG), Dec 05
- World Vision Australia (WVA); End of Project Review Report – Phase One, Rural Palestinian Women's Vocational and Income Generation Project (Gaza Strip), Dec 03
- WVA; Kariobangi HIV Project (Kenya), Jun 04
- WVA; Tamanna Child Labour Project Evaluation Report (India), May 02

b)   Read and review the following background documents:
- AusAID-NGO Cooperation Program (ANCP) Guiding Principles, AusAID
- List of ANCP Projects and ADPlans for each NGO (03-04, 04-05, 05-06)
- Accreditation documentation - Agency Profile Guidelines and Criteria
- Peter Ellis and Mike Crooke, "A Review for AusAID Accreditation Process and Systems," 2004

Each of the following broad tasks will be undertaken during the course of the metaevaluation:

Objective A: Analyse the NGO evaluations listed in the ToR to determine the quality of a cross-section of ANCP activity evaluations 2003-2005
- An examination of the evaluation methodology/frameworks
- An examination of the clarity and utility of the evaluations
- An examination of the risk management procedures noted in the evaluations

Objective B: Analyse the NGO evaluations listed in the ToR to determine the validity of the ANCP activity evaluation results and recommendations
- An examination of the level to which conclusions and recommendations are evidence-based (as opposed to speculative) or logically defensible
- An examination of lessons learnt

Objective C: Analyse the documents given to determine how broadly and consistently evaluation funds are being utilised by accredited NGOs.
- Examine if NGOs are consistently undertaking evaluations utilising ANCP funding
- An examination of which NGOs (and what percentage of NGOs) are undertaking evaluations utilising ANCP funding, and in which programming sectors
- Identification of any common themes (across NGOs and across sectors)
- Identify any lessons learned for AusAID