

NZAID 2007 Evaluations and Reviews
Annual Report on Quality

NZAID

August 2008

NZAID 2007 Evaluations and Reviews

Annual Report on Quality

Prepared by Miranda Cahn
Evaluation Advisor
Strategy, Advisory and Evaluation Group

NZAID

Wellington

August 2008

Contents

Contents.....	i
Abbreviations and Acronyms.....	ii
Executive Summary	iii
1. Introduction.....	1
2. List of evaluation and review reports included in the analysis	2
3. Methodology of this study.....	3
4. Analysis of the quality of the evaluation and review TOR.....	4
4.1 Quality of the TOR: questions asked	4
4.2 Quality of the TOR: findings of the study	4
5. Analysis of quality of Evaluation and Review Reports.....	9
5.1 Quality of the evaluations and reviews: questions asked.....	9
5.2 Quality of the evaluations and reviews: findings of the study	9
6. 2007 evaluations and reviews: areas for improvement	16
7. Trends in quality of NZAID evaluations/reviews	18
8. Conclusion.....	19
9. References used in this report.....	21
Appendix 1 Terms of Reference (TOR) for this study.....	22
Appendix 2 Schedule used for assessing quality of TORs	25
Appendix 3 Schedule used for assessing quality of evaluations and reviews ..	26
Appendix 3 Examples of good practice in the TOR and reports.....	29

Abbreviations and Acronyms

DAC	Development Assistance Committee (OECD)
ECOP	Evaluation Community of Practice (NZAID)
ERC	Evaluation and Research Committee (NZAID)
ERRC	Evaluation and Review Reference Collection (NZAID)
NZAID	New Zealand Agency for International Development
OECD	Organisation for Economic Cooperation and Development
SAEG	Strategy, Advisory and Evaluation Group (NZAID)
TOR	Terms of Reference

Executive Summary

Introduction

The New Zealand Agency for International Development (NZ Aid) Evaluation Policy Statement commits NZ Aid to increasing and improving the focus and use of both external and internal evaluative activity. Annually since 2004 NZ Aid has undertaken a desk study of the evaluations and reviews completed by NZ Aid during that calendar year. The 2007 study has been divided into two separate studies. The first part of the 2007 study focuses on the development learning and key issues from the evaluations and reviews, and the extent to which the conclusions and recommendations from earlier evaluations and reviews were implemented. This, the second part of the 2007 study, identifies key areas in the 2007 evaluations and reviews where the quality of evaluation/review Terms of Reference (TOR), evaluation and review practice, and reporting, could be improved. 18 reports were included in the 2007 quality study.

The Development Assistance Committee (DAC) Evaluation Quality Standards and the NZ Aid Evaluation Policy Statement guided the analysis in this study. The DAC Quality Standards incorporate guidelines on what constitutes 'good practice' in development evaluations and specifies five criteria for evaluating development activities: relevance, efficiency, effectiveness, impact and sustainability. The NZ Aid Policy Statement describes five principles by which evaluation should be guided: partnership, independence, participation, transparency and capacity building.

Study questions

The overall questions this study seeks to answer are:

1. Are the TOR for the evaluations and reviews of good quality?
2. Are the evaluations and reviews commissioned by NZ Aid of good quality?
3. What are the key quality aspects that should be improved in future evaluations and reviews?

Study methodology

The 2007 TOR were analysed for quality using the NZ Aid Guideline for Developing Terms of Reference for Reviews and Evaluations as a basis for setting quality criteria. Then the evaluation and review reports were analysed for quality using the DAC Quality Standards, DAC criteria and guiding principles of the NZ Aid Evaluation Policy Statement as reference. Areas for improvement in the 2007 evaluations and reviews were identified and compared with findings of

previous annual studies of evaluations (David 2005, Quinn 2006 and Clark 2007).

Quality of 2007 TOR

The TOR lacked quality in a number of key aspects. The way in which objectives and evaluation questions were presented was often not satisfactory. The description of methodology was either inadequate, or overly prescriptive, and the TOR generally did not present the methodology principles on which the evaluation should be based. Evaluation of value for money or cost effectiveness was not included in the objectives of most TOR, and reporting requirements for the evaluation or review were often not described well. Few TOR listed the important documents that should be referred to by the evaluators/reviewers. The study notes a number of improvements that could be made to the NZAID Guideline on Developing Terms of Reference for Reviews and Evaluations that would help to overcome some of the TOR quality issues.

Quality of 2007 evaluation and review reports

The study identified four reviews and evaluations (out of 18) where the overall quality was assessed as 'satisfactory or good in many respects'. Some of these reports were of a very high standard. Nine reviews and evaluations were assessed as 'satisfactory in some important respects' and five were assessed as 'not satisfactory in many respects'.

Most evaluation/review reports did not address the DAC criteria or the NZAID guiding evaluation principles, and the contextual and background sections often excluded the wider strategic and socio-economic contexts. Objectives and evaluation questions were often not well formulated.

Methodology was a major weakness in many of the evaluations and reviews. Few reports explained how appropriate stakeholders were selected, or how the evaluations/reviews ensured the findings were ethical and reliable. The methodology was often not described in enough detail, and did not include an explanation of why the chosen methods of information finding were used or how information gathered was analysed. Few of the evaluations or reviews were participatory. While for some a participatory approach was not appropriate, some would have benefited from using more participatory methods of information gathering. In some of the evaluations/reviews that said they were participatory, the methodology did not clearly describe how it was participatory. A few evaluations/reviews included analyses of impact, but these were generally not done well. Base-line data was not used in any of the evaluations/reviews in a systematic way when analysing impact.

Cross cutting and mainstreamed issues were generally not integrated well, and most reports did not address poverty elimination (or say why poverty elimination was not addressed). The reports generally did not assess value for money of the development activity being evaluated, or did not do so well. Finally, many of

the reports did not meet a satisfactory standard in terms of clarity, sequencing or completeness.

Areas for improvement

The following improvements would enhance NZAID commissioned reviews and evaluations:

1. Terms of Reference for evaluations and reviews
 - Follow more closely the Guideline for Developing Terms of Reference for Evaluations
 - Develop clearer objectives and evaluation questions
 - Develop less prescriptive, and more principles orientated descriptions of methodology. Include feedback of results to stakeholders as a requirement.
 - Include 'value for money' as one aspect to be evaluated
 - Better describe reporting requirements in TOR, including all types of reporting (such as workshops etc), feedback to stakeholders, and a timeline
 - Include reference to DAC Evaluation Quality Standards and NZAID Evaluation Policy Statement
2. Evaluations and reviews (and the resulting reports)
 - Address appropriate DAC criteria for evaluations and say why or why not criteria were included in the evaluation/review.
 - Be guided by the NZAID principles guiding evaluation activity stated in the NZAID Policy Statement, and describe this in the reports
 - Improve the contextual and background sections of the report
 - Improve the quality of objectives and evaluation/review questions and modify these from the TOR if this is appropriate
 - Improve the methodology of the evaluations/reviews and describe the methodology in more detail, including participatory approaches where appropriate.
 - Better integrate mainstreamed and crosscutting issues into the evaluations and reviews
 - Describe how the development activity being evaluated did or did not address poverty elimination, and address the effectiveness of the development activity in terms of poverty elimination if appropriate
 - Include, and improve the way in which value for money of the development activity being evaluated/reviewed is assessed
 - Improve the clarity, sequencing and completeness of the reports
3. Improvements to the NZAID Guideline on Developing Terms of Reference for Reviews and Evaluations

There was a correlation between the quality of TOR and the quality of evaluations and reviews. Improving TOR could positively impact on the quality of evaluations and reviews. Two other areas where improvement in quality would perhaps have the most impact on the quality of the evaluation and review reports are the formulation of objectives and evaluation questions, and methodology.

1. Introduction

The New Zealand Agency for International Development (NZAID) Evaluation Policy Statement notes that evaluation of NZAID development is critical for accountability, learning and improvement. Evaluation is defined in the Evaluation Policy Statement as “*an assessment, as systematic and objective as possible, of a planned, on-going, or completed development intervention*”. The term ‘review’ is used to describe evaluative activity taking place at key points during the lifetime of an activity.

The NZAID Evaluation Policy Statement commits NZAID to increasing and improving the focus and use of both external and internal evaluative activity. Annually since 2004 NZAID has undertaken a desk study of the evaluations and reviews completed by NZAID during that calendar year. The annual evaluation and review study reports are presented to the NZAID Evaluation and Research Committee (ERC) and disseminated within NZAID through various mechanisms including the Evaluation Community of Practice (ECOP), team meetings, Strategy, Advisory and Evaluation Group (SAEG) advisor engagement, the Intranet and the NZAID Evaluation and Review Reference Collection (ERRC).

The 2007 study has been undertaken in two separate studies, the first study¹ focuses on the development learning and key issues from the evaluations and reviews, and the extent to which the conclusions and recommendations from earlier evaluations and reviews were implemented. The report from the first 2007 study (*NZAID Annual Report of Development Practice Themes*) will be published on the NZAID website.

This, the second 2007 study of NZAID evaluations and reviews, aims to identify key areas where the quality of evaluation/review Terms of Reference (TOR), evaluation and review practice, and reporting, could be improved. The study encompasses all reviews and evaluations commissioned by NZAID, including studies commissioned jointly with other agencies, and completed during the 2007 calendar year. TOR for this study are in Appendix 1.

The Development Assistance Committee (DAC) Evaluation Quality Standards and the NZAID Evaluation Policy Statement guided the analysis in this study. The DAC Evaluation Quality Standards incorporate guidelines on the ‘key pillars’ needed for a quality evaluation process and product, and specify five evaluation criteria for evaluating development assistance: relevance, efficiency,

¹ Report title: *NZAID Annual Report of Development Practice Themes*, June 2008

effectiveness, impact and sustainability. Not all these criteria are appropriate for all the evaluations and reviews, and furthermore development partners may have other evaluation criteria which they consider important. However, the evaluation and review reports should clearly indicate which of these criteria, and which other criteria, are addressed in their evaluations/reviews, and why.

The NZAID Evaluation Policy Statement specifies five principles on which NZAID commissioned evaluation and reviews will be guided: partnership, independence, participation, transparency, and capacity building. Each NZAID commissioned evaluation or review should provide clear evidence that evaluations and reviews were guided by these principles.

The overall questions that this second study seeks to answer are:

- Are the TOR for the evaluations and reviews of good quality?
- Are the evaluations and reviews commissioned by NZAID of good quality?
- What are the key quality aspects that should be improved in future evaluations and reviews?

2. List of evaluation and review reports included in the analysis

For each evaluation or review, with a few exceptions, the TOR for the evaluations and reviews, the evaluation/review reports, and the submissions to the ERC on the evaluations/reviews were available.

	Evaluation/review	Comments
1	NZ Red Cross International Delegates Programme	
2	Volunteer Service Abroad	
3	Solomon Islands Primary School Storage Infrastructure Project	
4	Solomon Islands Primary School Operation Grants	
5	Cook Islands Country Strategy	
6	Building an Infrastructure for Research in the Cook Islands	ERC submission not available
7	Cook Islands ICT and STTA Awards Scheme	
8	NZDS and STTA Award Schemes in PNG	
9	Small Projects Scheme for Vanuatu	
10	Samoa NGO Support Fund	
11	Vanuatu Correctional Services Project	Report analysed was a draft report
12	Student Centred Teaching: Wan Smol Bag	
13	Pacific Health Research Council	
14	Pacific Invasives Initiative	ERC submission not available
15	Himalayan Trust	
16	Philippines National Ecotourism Programme	
17	Mekong HRD	
18	PNG Country Programme Strategy	ERC submission not available TOR not available

3. Methodology of this study

In order to analyse the 2007 evaluations and reviews for quality, first the criteria for analysing and assessing the TOR and the reports were established, and schedules for recording comments and assessments on the quality aspects of the TOR and the reports were developed (Appendices 2 and 3).

The TOR were analysed first. Using the schedule in Appendix 2 a quality assessment was made for each important aspect identified (with comments). Each evaluation/review report was then read, referring frequently to the corresponding TOR. The same process was used as for the quality assessment of the TOR, using the schedule in Appendix 3, and an overall quality assessment was given to each evaluation/review report. As the TOR and the evaluation/review reports were read and analysed for quality, a separate list of key issues and observations was compiled. This list, together with the quality assessments and comments in the schedules, formed the basis of the findings of the analysis of the quality of the 2007 evaluations and reviews in Sections 4 and 5. The submissions to the ERC for each evaluation/review were also read and any relevant points added to the overall analysis of quality.

Trends in the quality of evaluations and reviews over recent years were then analysed by referring to previous annual studies of the evaluations and reviews commissioned during each calendar year.

It should be noted that while it is the evaluations and reviews that are being analysed, the study is based solely on the written TOR and reports provided. This could be considered a limitation to the study as the quality of evaluations and reviews is dependant on factors that are sometimes not reported in the reviews and evaluations, such as poor design of the original project, illness of the evaluator, unavailability of stakeholders or other unavoidable factors. Furthermore, in some cases a commendable process (e.g. appropriate and robust methodology) may not have been comprehensively reported. However, the main purpose of this study is not to judge the quality of evaluations and reviews but rather to serve as a way of learning and improving. Despite the limitations of the study, valuable lessons have been learned.

4. Analysis of the quality of the evaluation and review TOR

TOR were available for all of the evaluations and reviews analysed except one. Not all reports included the TOR (six out of the 18 evaluation/review reports did not include TOR in the report), but TOR were available as part of the Evaluation Committee submissions.

4.1 Quality of the TOR: questions asked

TOR were analysed for quality against the NZAID Guidelines on Developing Terms of Reference for Reviews and Evaluations. The following questions were asked in terms of quality of the TOR:

- **Background and contextual information:** Did the TOR provide adequate background information –the history, the purpose, aims and objectives of the programme or project being evaluated/reviewed, the organisational context in which the evaluation will occur and the main stakeholders involved?
- **Purpose, scope and objectives:** Was the purpose of the evaluation or review clearly described – the rationale for conducting it, to whom results will be reported, how findings will be used? Was the scope of the evaluation or review described adequately? Were the objectives of the evaluation, and the questions to ask of the evaluation/review clearly articulated, and were they related to the purpose of the evaluation/review?
- **Methodology:** Was the approach to the evaluation methodology described well without being overly prescriptive?
- Did the TOR include questions about **cost effectiveness/value for money**?
- Was **management** of the evaluation described adequately, and the **composition of the evaluation team** described?
- Were **reporting** requirements described adequately, and where reports would be published? Was the expected format of reports described?
- Was evaluation/review **follow up** described?
- Were references provided for written material to be used in the evaluation/review?
- Did the TOR refer to the **DAC Evaluation Criteria** and the **NZAID Evaluation principles**

4.2 Quality of the TOR: findings of the study

Most of the TOR described the history of the programme or project being reviewed/evaluated well, and all described either the goals, objectives (or in one case activities) of the project. However, only six TOR clearly stated the main stakeholders in this section of the report.

All of the reports explained the purpose of the evaluation or review and the rationale for it, most did this well. However only half of the reports clearly stated

who the evaluation was aimed at, and furthermore only half the TOR specifically mentioned why the review was to be at this time.

Only five of the TOR fully and specifically described the scope of the evaluation or review including geographic area, timing, and activities, in a separate section. Some TOR listed tasks in a separate section, and while this was sometimes overly prescriptive, it did determine the scope of the activities. Only one TOR specifically mentioned areas that were outside the scope of the evaluation.

Six of the TOR stated the objectives of the evaluation/review, and the evaluation questions to be answered well, and in a manner that could provide excellent guidance in the evaluation/reviews and the reporting of them. In these TOR there was a good link between the evaluation/review purpose, objectives and questions. Other TOR listed objectives too briefly with no questions to be answered, or listed objectives and questions separately with an unclear relationship between them. In some the objectives were too broad, and in others objectives and/or questions to be answered were muddled and did not relate directly to the purpose of the evaluation. Sometimes tasks were included under the objectives, which leaves no room for the evaluator or reviewer, or other stakeholders, to decide how best to meet objectives, or answer evaluation questions.

The methodology section was arguably the weakest section of the TOR. The reports handled the methodology in a range of ways. Some of the reports only listed tasks in the methodology section, and one report combined a methodology section and a management section (very brief) and then listed tasks. One report included the methodology under the 'scope' section and was presented as a list of tasks. One report describes the methodology under the objectives – this description is very task orientated.

It clearly states in the NZAID Guideline on Developing Terms of Reference for Reviews and Evaluations that prescriptions of methodologies should be avoided and that the job of the TOR is to focus on the principles underpinning the evaluation/review and mechanisms to ensure 'an accountable and transparent process that builds local capacity'. While three of the TOR mentioned (one very vaguely) the core values that should underpin the evaluation, none of the methodologies focussed on approach and principles, and only one TOR addressed issues of accountability, transparency or approaches to ensure reliability of information by asking for triangulation of results.

Five of the TOR stated that the consultant/contractor was to develop a work plan and/or appropriate methods, or refine the 'draft design', but in none of the TOR was the work plan included as an output (or separate report) or as a milestone for payment.

A participatory approach to data collection (not for the whole evaluation/review) was suggested in five TOR, but only two of these explained how a participatory approach to data collection could be achieved. In the other three TOR the word 'participatory' was used in a very vague way.

One very important part of methodology is to describe who (what stakeholders) should be involved in the evaluation or review. Involving the appropriate stakeholders is important to the reliability of the information collected and the overall quality of the evaluation or review. This was poorly described in the TOR. One TOR said 'ensure key stakeholders are included'. Three TOR listed the stakeholders in a separate annex. Several mentioned stakeholders in lists of tasks or activities, and some listed major stakeholders in the background information. None of the TOR asked consultants or contractors to conduct a stakeholder analysis.

Feeding back and discussing evaluation results as part of the evaluation/review, is an important way of verifying findings, insuring all stakeholders are included, and ensuring accuracy. Nine (around half) of the TOR clearly spelled out how findings could be fed back to stakeholders for discussion or input before the results were finalised. One TOR says that the evaluation is for 'the community' but does not say how it will be fed back to the community.

Eight (almost half) of the TOR asked for some analysis of the cost effectiveness or value for money of the project or programme being evaluated/reviewed. Two of the TOR stated how much had been spent on the projects/programmes, but one of these did not then ask the evaluation team to assess value for money.

A section on 'management' was either missing from TOR or did not cover all the aspects listed in the NZAID Guideline on Developing Terms of Reference for Reviews and Evaluations. Only two TOR specifically stated who was commissioning the evaluation or review, a few mentioned how the draft reports and/or support to the team would be handled. A few specifically mentioned responsibilities of evaluation/review teams and or the responsibility of steering groups/committees for the evaluation.

Almost all of the TOR listed the composition of the evaluation team and the skills or competencies required. Three did not mention the composition of the team. One mentioned the composition of the team under the heading of 'management'.

In general reporting requirements were not described in sufficient detail in the TOR to give a clear idea of what was expected, and not all the TOR included specific times when the reports were to be submitted, or included other methods of reporting (e.g. workshops) although these were sometimes mentioned elsewhere in the TOR. Only two of the TOR mentioned that summaries of the evaluation/reviews would be available on the NZAID website, and full reviews

would be available if requested. Only two TOR mentioned that reports should adhere to the NZAID Guidelines on the Structure of Evaluation and Review Reports. Few TOR mentioned specific headings to use. Three of the TOR specifically mentioned the length of the report but in one case it was only 12 pages which seems rather short for the evaluation in question. Only one report mentioned whether reports should be in electronic or hard copy or both. Some TOR described 'outputs' rather than 'reporting'. This seemed sensible as in some of the TOR workshops and or presentations were required in addition to reports.

Around half of the TOR specified that the findings would be fed back to stakeholders (see section on methodology above). Usually this was asked for in the form of a draft report and in a few cases a workshop or presentation was specified. However, there was little in the TOR about how the outcomes from evaluations would be followed up.

Only four of the TOR clearly listed the written material to be used in the evaluation/review. The DAC Evaluation Quality Standards or the NZAID Evaluation Policy were not mentioned in any of the TOR as references to which the consultants/contractors need to be familiar with. Only one TOR mentioned the NZAID Guidelines on the Structure of Evaluation and Review Reports.

The TOR rarely specified or referred to the DAC criteria² or NZAID evaluation principles that should guide NZAID evaluations/reviews.

Table 1 (next page) summarises the quality assessments of different aspects of the TOR.

² It is noted that the inclusion of DAC criteria was not explicitly mentioned in the NZAID Guideline on Developing Terms of Reference for Reviews and Evaluations.

Table 1 Summary of assessment of quality aspects of evaluation and review TOR

Report #	Addressed DAC criteria for evaluation	Addressed DAC Quality standards	Addressed NZAID principles guiding evaluation	Background and context clearly described	Rational and purpose clearly described	Scope clearly explained	Objectives clearly explained	Methodology	Cost effectiveness asked for	Management and team composition described	Reporting requirements described (including workshops etc)
1	*	***	**	***	***	***	**	**	***	***	***
2	*	*	*	**	***	***	**	*	*	***	*
3	*	*	*	***	**	***	**	*	***	**	**
4	**	*	*	***	***	***	***	*	*	***	**
5	*	*	*	***	*	***	**	*	n/a	*	**
6	*	*	*	***	**	*	***	*	***	**	**
7	*	*	*	***	*	*	***	*	*	**	**
8	*	*	*	**	***	***	***	*	*	*	*
9	**	*	*	***	**	*	**	*	*	**	**
10	*	*	*	***	**	*	**	*	*	**	**
11	*	*	*	**	**	*	*	**	***	**	*
12	*	*	*	*	***	***	*	*	*	**	**
13	*	*	*	**	***	**	***	*	*	**	***
14	*	*	**	***	***	*	***	*	***	**	**
15	*	*	*	**	**	**	**	*	**	**	***
16	**	*	**	***	*	**	*	**	***	**	**
17	***	*	*	**	**	*	*	**	***	***	**
18	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a

Quality criteria

- * quality criteria NOT met in many or most respects
- ** quality criteria met to a satisfactory level in some respects
- *** quality criteria met to a satisfactory or high level in most respects.

5. Analysis of quality of Evaluation and Review Reports

5.1 Quality of the evaluations and reviews: questions asked

The questions asked of each of the evaluations and reviews analysed were:

- Did the evaluation/review report **meet the TOR**, if not did it explained why not?
- Is the **background** to the development activity being evaluated/reviewed, and is the policy, institutional and socio-political **context** of the evaluation clearly described?
- Are the **rational, purpose, objectives and scope** of the evaluation clearly explained?
- Is the **methodology** used sound, provides reliable information, in line with NZAID's core values, and clearly explained? Does the methodology include a description of stakeholder contribution, information sources, issues of transparency, and explanation of how reliability and accuracy of information was ensured? Are mainstreamed and cross cutting issues considered?
- Is there evidence that **evaluation ethics** are incorporated into the evaluation/review including independence, cultural and gender sensitivity, and are any constraints to or disagreement between the evaluation/review team declared?
- **Utility and relevance** – are the findings and lessons learned relevant, clearly articulated, policy/operationally orientated and useful?
- Did the evaluation/review reports address the issue of **value for money of the projects and programmes** well?
- **Completeness, and quality of the report** – is there a logical flow to the report, clear reporting of information sources, good explanations, appropriate style and tone, a clear and comprehensive executive summary, and clear recommendations that are directly related to the findings of the evaluation/review.
- Has the evaluation or review clearly articulated which of the **DAC criteria for evaluation** have been addressed?
- did the evaluation/review report clearly indicate that the evaluation/review had been guided by the **NZAID evaluation principles** as they are stated in the NZAID Evaluation Policy Statement?

5.2 Quality of the evaluations and reviews: findings of the study

Most of the reports did meet the TOR. In most cases the objectives as mentioned in the TOR formed the basis for structuring the report and thus the evaluations and reviews addressed the questions asked. However, when the objectives were not clear in the TOR, the reports were muddled and difficult to follow. In a few cases some of the objectives and/or questions in the TOR were

ignored. In one report a reason for this was provided but disputed by the ERC submission.

In most reports the background to the development activity being evaluated/reviewed was similar or the same as it was in the TOR. In general, where the background to the develop activity had been described well in the TOR, then it was also described well in the evaluation/review.

In general, descriptions of the institutional context of the development activity being evaluated were done well. However, socio-economic context was rarely described, and this constrained the understanding of how the activity being evaluated or reviewed aligned (or did not align) with the NZAID overarching goal of poverty elimination.

Usefulness of the evaluation to NZAID is also constrained if evaluations/reviews do not address the wider policy and strategic context within which the project/programme exists. This was mentioned by David (2005) as an area lacking in the reports that she analysed. A description of the wider strategic and policy background was lacking in around half of the 2007 evaluation and review reports. Few of those that described the policy and strategic background did so well.

For most of the reports the rationale for the review or evaluation, and its purpose were clear. However, few specifically mentioned who had commissioned the review (in most cases this was NZAID, but occasionally other organisations and agencies were also involved). Most reports mentioned how the evaluation or review would be used, but detail was generally lacking.

Objectives listed in the reports closely followed those specified in the TOR. Where objectives were poorly formulated in the TOR, none of the reports modified the objectives. Rather, one report ignored one objective, others barely addressed unrealistic objectives and others tried to prepare a report around poorly formulated objectives. Evaluation question are likely to evolve as evaluations or reviews progress, and stakeholders raise issues that the TOR had not expected or addressed. Evaluators and reviewers should, in some cases, have modified the objectives and/or the evaluation questions in order to accommodate unexpected issues (explaining why they had done so).

The scope of the evaluations and reviews was rarely specified according to the standards of the DAC Evaluation Quality Standards which mention that scope should cover geographic specifications, time, and activities, and also the criteria that are to be included in the evaluation. Aspects that fall outside the scope of the evaluation/review should also be described. While the scope of the evaluation/review could usually be understood once the entire report had been read, it would have been useful to specify this in the introductory sections of the reports.

The methodology of evaluations and reviews is critical to ensure a useful and relevant evaluation or review that is guided by the NZAID principles described in the NZAID Evaluation Policy Statement (partnership, transparency, independence, participation (where appropriate), and capacity building). A sound methodology ensures that crosscutting and mainstreamed issues are considered as in the NZAID Screening Guide for Mainstreamed and Other Cross Cutting Issues, that stakeholders' views and 'voices' are included, that stakeholders 'own' the results of the evaluation or review, and that the evaluation or review serves the purpose for which it was designed.

In one third of the evaluation/review reports the methodologies described did not meet the quality criteria in many (or most) respects. In a further one third, the methodologies described met the quality criteria in only some respects. In only three of the evaluations/reviews did the methodology meet the quality criteria in most respects. Although several of the TORs asked consultants/contractors to develop appropriate methodologies, only in two reports was there evidence that this had been done well. Constraints were rarely listed in the methodology section.

Identification of stakeholders: In most cases the stakeholders that were interviewed or consulted were listed. However, none of the evaluations or reviews conducted a specific stakeholder analysis to determine (for example) who were the primary and secondary stakeholders, what were their interests in the development activity and the evaluation/review, what was their level of influence and/or power, how these aspects may have biased or influenced the outcome of the evaluation or review, and how they would be engaged in the evaluation. This would have been appropriate and useful in many of the evaluations and reviews.

Explanation of methodology, and data analysis: In many of the reports the methodology was poorly explained, and did not clearly explain the processes used to engage stakeholders and answer the evaluation questions. Where field work was required, interviews and meetings were often the only method of information gathering used. No explanation was provided as to why these were considered the most appropriate methods. Rarely were the different methods of engaging stakeholders specified against the different questions and/or objectives of the evaluation. However, against the trend, one report presented a table which included the evaluation questions, and the methods used to find information to answer each question. This table provided excellent information on the methods of data collection used.

The way in which information and data was analysed and synthesised was only explained satisfactorily in one evaluation/review report. Most of the reports did not append any list of questions that were asked in interviews and meetings. Usually questionnaires used in surveys were appended, but in one report the

questionnaire used in the evaluation/review was not included in the appendices so it was impossible to know what had been asked.

Reliability of information: In most of the evaluations/reviews the reliability of information gathered for the evaluations and reviews was questionable. Around one quarter of the reports stated how evaluators/reviewers had ensured reliability of information they had gathered. One evaluator had attempted to triangulate data by using three different data sources to answer on evaluation question, and several reports described using follow-up in-depth qualitative interviews or field visits to verify quantitative survey results or questionnaire data. Where meetings and/or informal interviews formed the basis for the evaluation or reviews there was seldom any evidence to provide the reader with an idea of how findings were derived, or who had provided what information. Only in two reports were direct quotes from the stakeholders included in order to highlight certain issues and illustrate the findings. Direct quotes provided the reader with a good sense of the tone of the interviews and resulting data.

Participatory approaches and methods: Participatory approaches and methods were not common in the 2007 evaluations and reviews. Participatory approaches are not appropriate for all evaluations and reviews, and where the development activity has not been participatory, it would normally be inappropriate to expect a participatory evaluation or review as described in NZAID Guideline : Participatory Evaluations. However, use of participatory methods could have enhanced the findings and reliability of some of the evaluations/reviews. Six TOR mentioned a 'participatory approach' or using participatory methods, but in two evaluations/reviews where a participatory approach was asked for in the TOR this was not implemented (in one case because of lack of time). One report where the TOR asked for participatory approach mentioned that a participatory approach had been used but there was no evidence to support this.

Five of the reports indicated that participatory approaches had been used. In three of these reports participation had been limited to inclusion of primary stakeholders, or using participatory methods such as semi-structured interviews, focus groups or workshops. Only two of the reports described in detail a participatory approach that the evaluation/review had used. These reports showed that a participatory approach to evaluations and reviews, where appropriate, and when done well, can greatly enhance ownership, and the overall quality and usefulness of the evaluation/review

Use of base-line data: Most reports did not refer to base-line data, or explain in a systematic way why base-line data was not being used. Even when impact was attempting to be assessed, the evaluation/ reviews generally did not attempt to compare the present situation with any base-line.

Most of the evaluations were useful and relevant. Even where all the objectives were not met, and the reports did not meet all or even most of the quality criteria, the evaluations and reviews delivered insightful and useful information that ensured that the evaluations and reviews could be used for the purpose they were intended. However, in two evaluations/reviews this was after the final reports had been returned for rewriting after the ERC submission, and in another evaluation/review the report provided little useful information. In one review, a desk study was commissioned for the review of a country strategy. The usefulness of the review was constrained by some of the critical reports not being available and it was questionable whether in fact a desk study was appropriate for this review.

Usefulness of evaluations and reviews to NZAID is reliant on the evaluations/reviews addressing mainstreamed and crosscutting issues, and the overarching goal of poverty elimination that are espoused by NZAID. Braun (2008) identifies areas within the 2007 evaluation/reviews where these are addressed. Only three of the reports analysed for quality specifically addressed poverty elimination in their reports.

In terms of crosscutting issues, gender equality was rarely specifically addressed. Five reports mentioned issues of gender but systematic and detailed gender disaggregated analysis of the effects or impacts of projects, or any other gender issues was generally lacking. Other issues such as human rights, environment, conflict prevention and peace building, and HIV/AIDS were mentioned in evaluation/reviews where the projects were directly related to that issue.

Only around one third of the TOR explicitly asked for an assessment of whether the project or programme was cost effective or providing value for money. The first 2007 study of evaluations and reviews (Braun, 2008) analysed the evaluations and reviews for their information on cost effectiveness. Braun (2007) found that several evaluations and reviews had addressed the cost effectiveness although it was not asked for in the TOR and in total she described the information from ten reviews. In most cases these evaluations /reviews reported whether objectives had been met and/or whether there had been positive impacts, assuming that if they had, then this represented good value for money spent. The quality of information about value for money of projects could have been improved in most of the evaluations and reviews.

The quality of the reporting varied considerably. The quality of the TOR had a big impact on the quality of the evaluation/review report. In most cases, where a TOR was of a high standard then the report reflected the comprehensive guidance provided by the TOR. The appendices (annexes) of the reports varied considerably. Some reports were inadequate in terms of the appendices included. Six did not include the TOR. One report that was a desk study did not

include a list of written material used. Three reports did not have a contents page and two reports did not name the author.

In most cases the language used in the reports was appropriate, and the reports were presented adequately or well. Some reports did not flow well, the sequencing of sections was illogical or there were key sections missing. The executive summaries were sometimes too long or too short. In a few cases recommendations were not directly related to the findings of the review or evaluation. Recommendations were not always clearly summarised or listed at the front of the report, or with the executive summary.

TOR rarely specified which of the DAC criteria for evaluation (efficiency, effectiveness, relevance, sustainability and impact) the evaluations and reviews were specifically to address. In the few instances where TOR did mention DAC criteria, then the reports also specified which of the DAC criteria were being addressed. Sometimes the reports used the words impact, or development effectiveness, or sustainability, but it was unclear whether the other criteria had been considered or whether this was just the rhetoric of the report (or TOR).

Only two of the reports specifically mentioned NZAID principles guiding evaluation (partnership, independence, participation, transparency and capacity building).

Table 2 (next page) summarises assessments for different aspects of quality of evaluation and review reports, and provides an overall assessment of the quality of the reports.

Table 2 Summary of assessment of evaluation and review reports quality

Report #	Addressed DAC criteria for evaluation	Addressed NZAID principles guiding evaluation	Eval/rev report met TOR	Background and context clearly described	Rational and purpose clearly described	Objectives clearly explained	Scope clearly explained	Methodology developed well	Evaluation ethics addressed	Ev/rev Findings etc useful and relevant	Reporting done well	Overall Assessment of quality
1	*	*	***	***	***	**	**	**	*	***	***	√√
2	*	***	***	**	***	**	***	***	***	***	**	√√
3	***	*	***	***	***	**	***	**	*	***	**	√√
4	*	*	***	***	***	***	***	***	*	***	***	√√√
5	*	*	***	***	***	**	***	*	*	*	*	√
6	**	*	***	***	***	***	***	***	**	***	***	√√√
7	*	**	*	***	***	***	***	*	*	*	*	√
8	***	***	***	***	***	***	***	*	*	***	**	√√
9	***	*	**	***	***	**	*	**	*	***	**	√√
10	*	*	*	**	***	***	*	*	*	**	*	√
11	*	*	**	**	**	*	*	*	*	**	**	√√
12	*	*	**	**	*	*	*	***	**	*	*	√
13	*	*	***	**	***	***	**	***	**	***	***	√√√
14	*	*	**	***	***	***	*	**	*	***	**	√√
15	*	*	*	**	**	*	**	*	*	**	*	√
16	***	**	***	**	**	**	**	***	**	***	***	√√√
17	**	*	***	***	***	*	*	***	*	***	***	√√
18	*	*	n/a	***	***	**	*	**	*	***	**	√√

Quality criteria

- * quality criteria NOT met in many respects
- ** quality criteria met to a satisfactory level in some respects
- *** quality criteria met to a satisfactory or high level in most respects.

Note that some quality criteria give more weight to the overall quality of the evaluation review report, and the overall assessment is therefore subjective.

Overall assessment of quality

- √ not satisfactory in many important respects
- √√ satisfactory in some important respects
- √√√ satisfactory or good in many respects

6. 2007 evaluations and reviews: areas for improvement

The findings of this study on quality of the 2007 evaluations and reviews (reported in Sections 4 and 5) indicate that the following improvements would most enhance NZAID commissioned reviews and evaluations:

1. Terms of Reference for evaluations and reviews

- Follow more closely the Guideline for Developing Terms of Reference for Evaluations
- Develop clearer objectives and evaluation questions
- Develop less prescriptive, and more principles orientated descriptions of methodology. Where consultants are to provide a work plan or separate methodology make this a specific requirement in the TOR and note it as a milestone. Ensure TOR require that a stakeholder analysis be completed. Include feedback of results to stakeholders is a requirement.
- Include 'value for money' as one aspect to be evaluated
- Better describe reporting requirements in TOR, including all types of reporting (such as workshops etc), feedback to stakeholders, and a timeline
- Include reference to DAC Evaluation Quality Standards and NZAID Evaluation Policy Statement

2. Evaluations and reviews, and the resulting reports

- Address appropriate DAC criteria for evaluations and say why or why not criteria were included in the evaluation/review.
- Be guided by the NZAID principles guiding evaluation activity stated in the NZAID Policy Statement, and describe this in the reports
- Improve the contextual and background sections of the report
- Improve the quality of objectives and evaluation/review questions and modify these from the TOR if this is appropriate
- Improve the methodology of the evaluations/reviews. Ensure all appropriate stakeholders are included and their voices are clearly heard, the findings are ethical and reliable, the methodology is clearly described (including methods used and how information was analysed), if the evaluation is participatory, the methodology clearly describes how it is participatory, and the process is truly participatory. Where it is appropriate to evaluate impact, improve the way in which this is done and use any available base-line data. Describe the methodology in more detail in the reports.
- Better integrate mainstreamed and crosscutting issues into the evaluations and reviews
- Describe how the development activity being evaluated did or did not address poverty elimination, and address poverty elimination when evaluating the effectiveness of the development activity if appropriate
- Include, and improve the way in which value for money of the development activity being evaluated/reviewed is assessed
- Improve the quality, and completeness of the reports

3 The analysis of the TOR highlighted some areas where the NZAID Guideline on Developing Terms of Reference for Reviews and Evaluations could be improved. These improvements may lead to improvement in future TOR. For example:

- Include instruction in the guideline to include in TOR which of the DAC criteria should be addressed in the evaluation.
- Rename the section on 'reporting requirements' in the guideline as this is confusing when expected outputs of the evaluation are other than reports (e.g. workshops conducted, feedback to stakeholders, work plan etc).
- Include explanation in the guideline that where a work plan is requested, and is expected to describe the methodology, the work plan needs to be included as an output in the TOR, and noted in the milestones. The methodology design should appear in the appendices of the report.
- The explanation in the guideline on how to prepare objectives in TOR is very brief and does not provide sufficient guidance. Examples would help.
- The guideline needs to explain that TOR should mention that the evaluation should conform to the DAC Quality Standards, and the principles in the NZAID policy statement.

7. Trends in quality of NZAID evaluations/reviews

There are a number of quality issues that have consistently been noted by the authors of the annual studies of evaluations and reviews (David 2005, Quinn 2006, Clark 2007), as well as in this report (Table 3, Page 21).

David (2005) and Quinn (2006) both note that many of the evaluations and reviews they analysed do not focus on impact. In this report it was found that where impact analysis was identified as appropriate it was not done well.

Inclusion of crosscutting issues has been identified as a quality issue in all the previous annual analyses and syntheses of evaluations and reviews. David (2005) mentioned that there was a 'frequent absence of gender analysis in the reports' (David 2005) while Quinn (2006) noted that many TOR and the majority of reports failed to identify or address gender issues. Clark (2007) noted the absence of cross-cutting issues (including gender) in the 2006 evaluations and reviews, and this report also identifies issues in this area.

A focus on poverty has also been lacking over the past four years. While, it is unrealistic to expect poverty elimination to be a focus of evaluations if poverty elimination has not been a focus in the design of the development activity, evaluations and reviews should address whether the development activity does or does address poverty elimination and if not why not.

David (2005) and Quinn (2006) report a lack of participatory process in the evaluations and reviews they analysed, while Quinn (2006) also mentioned there was a lack of understanding of what was expected of a participatory approach. Clark (2007) noted that some reports indicated a participatory approach had been used but few reported how they did so. This 2007 study also noted weaknesses with use of participatory approaches in the few evaluations/reviews where participation was appropriate and was an aspect of the evaluation or review.

Weakness in methodology was noted by Clark (2007) and Quinn (2005). This study also found serious issues with the methodologies used and the way in which they were described. Weakness in reporting was noted by Quinn (2006), Clark (2007) and also this study.

Quality of TOR was identified as an issue by David (2005), and was a focus of the 2007 report (Clark 2007) and this report.

Analyses of the quality of 2006 and 2007 NZAID evaluations/reviews have identified shortcomings in areas such as ethics, independence and reliability of the evaluations/reviews, and in the way that evaluations and reviews address the DAC criteria, and the guiding principles in the NZAID Evaluation Statement.

Table 3 notes quality issues identified in the annual studies of evaluations and reviews reports over the past four years.

Table 3 Evaluation/review quality issues highlighted in annual studies of NZAID evaluations and reviews

Issue	David (2005)	Quinn (2006)	Clark (2007)	Cahn (2008) Braun 2008)
Quality of TOR	+		+++	+++
Clarity of purposes and objectives			+++	++
Methodology		+++	++	+++
Participatory processes	+++	+++	++	+++
Focus on poverty	+++		++	+++
Assessment of value for money	+++			+++
Focus on impact	+++	+++		+
Inclusion of gender	+++	+++		
Inclusion of crosscutting and mainstreamed issues (including gender)			+++	+++
Wider issues drawn out	+++			
Reporting issues		+++	+++	+++
Ethics (including independence)			++	++
Usefulness of the evaluation/relevance of results			++	++
DAC criteria addressed			+++	++
Guided by NZAID Evaluation			+++	++

+Mentioned but not considered to be a serious issue

++Mentioned as an issue

+++Emphasis placed on this issue

8. Conclusion

All of the 2007 TOR and the evaluation and review reports had room for improvement. However, this study identified four reviews and evaluations (out of 18) where the overall quality was assessed as 'satisfactory or good in many respects'. This was the highest assessment 'grading', and some of these evaluations/reviews were of a very high quality. Nine evaluations or reviews were assessed as 'satisfactory in some important respects'. Five evaluations and reviews were assessed as 'quality criteria NOT met in many respects'.

The major limitations of the evaluations and reviews that this study identified were:

- Quality of TOR
- Formulation of objectives and evaluation questions
- Methodology (including weakness in participatory approaches and methods)
- Assessment of value for money
- Focus on poverty
- Inclusion of crosscutting and mainstreamed issues
- Reporting issues

The study identified a number of key areas where improvements to TOR and to evaluation and review practices and reporting would greatly enhance the quality of NZAID evaluations and reviews (summary in Section 6). The study identified a correlation between the quality of the some of TOR and the quality of some of the evaluation/review reports. Thus if the quality of the TOR can be improved, this may result in improving the quality of the evaluations and reviews and their reports. The study noted that improvement to the NZAID Guideline on Developing Terms of Reference for Evaluations could help to improve future TOR. Two other areas where improvement of quality could have considerable impact on the quality of NZAID evaluations and reviews are formulation of objectives and evaluation questions, and methodology.

Familiarising evaluators and reviewers with NZAID evaluation guiding principles and DAC Evaluation Quality Standards is also likely to have a favourable impact on the way in which the evaluations and reviews are conducted.

Some of the areas of potential improvement identified in this study have also been identified in previous annual studies of reviews and evaluations, namely the quality of TOR, a focus on poverty, inclusion of crosscutting and mainstreamed issues (especially gender), clarity of purpose and objectives, methodology, reporting, ethics, and the way in which the evaluations address NZAID evaluation principles and DAC criteria.

The NZAID Evaluation Policy Statement commits NZAID to increasing and improving the focus and use of both external and internal evaluative activity. In order to do this, the key quality issues identified in the annual studies of evaluations and reviews need to be further addressed within the organisation, reasons for lack of improvement identified, and processes for ensuring that improvement does occur developed.

9. References used in this report

NZAID Evaluation and Research Committee submissions

NZAID Guideline: Participatory Evaluations

NZAID Screening Guide for Mainstreamed and other Cross-cutting Issues

NZAID Evaluation Policy Statement

NZAID Guideline on Developing Terms of Reference for Reviews and Evaluations

NZAID Guidelines on the Structure of Evaluation and Review Reports

Braun, A. 2008 NZAID Annual Report of Development Practice Themes. Unpublished report.

Clark K. 2007. Annual Summary Report: NZAID Reviews and Evaluations Undertaken In 2006. Unpublished report.

DAC Evaluation Quality Standards. DAC Evaluation Network. OECD

David R. 2005. Looking Back & Learning: A summary of and lessons from NZAID Evaluation Reports June 2002 – December 2004. Unpublished report.

Quinn M. 2006. Summary Report NZAID Reviews and Evaluations January – December 2005. Unpublished report.

Appendix 1 Terms of Reference (TOR) for this study

Terms of Reference 2007 Annual Report on the quality of NZAID Evaluations and Reviews

Background

The NZAID Evaluation Policy Statement defines evaluation as “*an assessment, as systematic and objective as possible, of a planned, on-going, or completed development intervention*” and notes that evaluation of NZAID development is critical for accountability, learning and improvement. The term ‘review’ is used to describe evaluative activity taking place at key points during the lifetime of an activity. The DAC guiding principles (partnership, independence, participation, transparency and capacity building), DAC evaluation criteria (relevance, effectiveness, efficiency, impact, sustainability) and DAC Quality Standards³ for the evaluation of development assistance complement the NZAID Evaluation Policy Statement and the related set of practice guidelines. The NZAID Evaluation Policy Statement commits NZAID to increasing and improving the focus and use of both external and internal evaluative activity, and in order to do this, is committed to improving its capacity to commission and undertake good quality evaluations.

Since 2004 NZAID has undertaken a desk-based study of the evaluations and reviews commissioned during each calendar year. The studies serve to firstly synthesise and summarise development learning from evaluations and reviews, and secondly to assess the quality of the studies, and highlight areas for further improvement. The 2007 study has been divided into two separate reports, the first report⁴ focuses on the development learning and key issues from the evaluations and reviews and the extent to which the conclusions and recommendations from earlier evaluations and reviews were implemented, and this second study focuses on the quality aspects of the evaluations and reviews, including the quality of the TOR prepared for them.

The reports are presented to NZAID’s Research and Evaluation Committee and disseminated within the agency through various mechanisms including the Evaluation Community of Practice, team meetings, SAEG Adviser engagement, the Intranet and the ERRC collection. The first report (*NZAID Annual Report of Development Practice Themes*) will also be published on NZAID’s website.

³ OECD-DAC. 2006. DAC Evaluation Quality Standards

⁴ Report title: *NZAID Annual Report of Development Practice Themes*, June 2008

Purpose

The purpose of the second study is to contribute to organisational learning in NZAID through an analysis of NZAID-commissioned evaluations and reviews to identify key areas where the quality of evaluation/review practice, quality of TOR, and reporting quality, could be improved.

Scope

The study will encompass all reviews and evaluations commissioned by NZAID, including studies commissioned jointly with other agencies, and completed during the 2007 calendar year.

Key Questions

With respect to evaluative activity quality

1. Are the terms of reference for the evaluations and reviews of good quality and consistent with the NZAID terms of reference guidelines?⁵ How could they be improved?
2. Are the evaluations and reviews commissioned by NZAID of good quality? Do they comply with NZAID Evaluation Policy, Guidelines and DAC Evaluation Standards?
3. What are the key quality aspects that should be improved in future evaluations and reviews to provide better accountability and learning, and lead to improvement of NZAID's development activities.

Process and Methodology

The Evaluation Advisor assigned to this analysis will:

- assess the 2007 evaluation and review reports against NZAID's Evaluation Policy and the DAC Evaluation Standards,
- prepare a draft report
- seek feedback from the Evaluation Team and ECOP and incorporate this into the report
- produce a final report.

Outputs

The report :

2007 Annual Report on the quality of NZAID Evaluations and Reviews will include the following sections:

⁵ NZAID Guidelines on Developing Terms of Reference for Reviews and Evaluations.

- A. Executive Summary
- B. Background/introduction
- C. List of evaluation and review reports included in the analysis
- D. Methodology, and basis of analysis
- E. Analysis of the quality of the evaluation and review TORs
- F. Analysis of the quality of the evaluation and review reports
- G. Terms of Reference
- H. References (e.g. see 'Other documents' subsection)
- I. Analytical appendices as necessary

The report (excluding contents, executive summary, acronyms, and appendices) should not exceed 10 pages.

Key Dates and Milestones

Date	Event	Milestone
13 August 2008	Task assigned to evaluation advisor	
27 August 2008	Draft report completed	Report submitted
3 September 2008	Final report completed	Peer-review feedback on the draft report incorporated and final report submitted.

Follow-up

The report findings will be discussed by the Evaluation Community of Practice and the report submitted to the Research and Evaluation Committee.

Information Sources and Key Documents

NZAID-commissioned Evaluation and Review Reports
 NZAID Research and Evaluation Committee Submissions
 Previous Evaluation and Review Summary reports (2004, 2005, 2006 and 2007).

Other key documents:

NZAID Evaluation Policy Statement
 DAC Evaluation Quality Standards
 NZAID Guideline: Structure of Review and Evaluation Reports
 NZAID Guideline: Developing Terms of Reference for Reviews and Evaluations
 NZAID Guideline: Participatory Evaluations
 NZAID Screening Guide for Mainstreamed and Other Cross-cutting Issues

NZAID Evaluation Strategy

Appendix 2 Schedule used for assessing quality of TORs

Schedule for TOR of NZAID 2007 Reviews and Evaluations

Note: This schedule was developed specifically for this assignment. It was guided by NZAID Guideline for Developing TOR for Reviews and Evaluations.

Title of Review/Evaluation:

QUALITY	Excellent	Satisfactory	Poor	Comment
Background and Context described well				
Rational and purpose				
Scope described				
Objectives of evaluation well written and relate to purpose				
Methodology – as per guideline				
Cost effectiveness/ value for money included				
Management described				
Team described				
Reporting described				
Evaluation follow-up				
Specified DAC Criteria				
Specified NZAID principles				

Notes:

Appendix 3 Schedule used for assessing quality of evaluations and reviews

Schedule for quality assessment of NZAID 2007 Reviews and Evaluations

Note: This schedule was developed specifically for this assignment. It was adapted from a checklist developed earlier by SAEG staff, and guided by DAC Quality Standards.

Title of Review/Evaluation:

QUALITY	Excellent	Satisfactory	Poor	Comment
Did report meet TOR				
Rational and Purpose				
Objectives				
Scope				
Background and Context				
Methodology				
Evaluation ethics				
Utility and relevance				
Reporting				
Value for Money				
Specified DAC Criteria				
Specified NZAID principles				
OVERALL				

Notes:

Quality Domains	Excellent	Satisfactory	Poor	Comments
	Did report meet TOR			
Background and context (policy and institutional and scio-cultural) The relevant policy objectives and principles have been set out The institutional context of the subject has been described Policy context described Socio-cultural context described The background and history of the programme/activity are described clearly. The implementation of the intervention is described in sufficient detail.				
Rational and Purpose Description of the rationale Description of purpose The purpose is in line with NZAID Evaluation Policy The purpose has been formulated and described in clear terms How the information will be used is included				
Scope defined Is scope clearly set out – geographic, activities, time period, etc. Evaluation criteria are clear. The parameters are clear – what’s included and what’s not				
Formulation of review / evaluation questions - Key questions stated - Sub-questions clearly link to key questions - Objectives linked to purpose				
Methodology - Key stakeholders are identified - Methodology is clearly explained, including any constraints. - The rationale for choice of methodology is explained. - Methods of data included - Methods of data analysis included.				

<p>Evaluation ethics Evaluation ethics included – independence, cultural and gender sensitivity, constraints declared, disagreements in team declared.</p>				
<p>Utility and Relevance The purpose of the review/evaluation is fulfilled The results are relevant to the intended use of the evaluation Recommendations are policy and/ or operationally orientated -Report is available in time to inform decisions</p>				
<p>Completeness and quality of the report</p> <ul style="list-style-type: none"> - Sections logical and logical flow - Information sources and any gaps in the information available have been reported - Differences between expected and actual results explained - Style and tone of report is appropriate - Sensitive material has been put into a “Confidential Annex” to the main report. - The report has a clear executive summary that provides an overview and includes the conclusions, recommendations and key learning. 				
<p>Value For Money The evaluation assignment has been implemented within the allotted time and resources The review/evaluation questions have been addressed The evaluation products are useful to NZAID and other key stakeholders</p>				
<p>DAC criteria addressed Have the DAC criteria been acknowledged and does the report specify which are being addressed? Have the criteria been properly addressed?</p>				

Appendix 3 Examples of good practice in the TOR and reports

Despite shortcomings described in the study, there were numerous examples in the 2007 TOR and evaluations and review reports of good practice. A few of those are described here.

1. Terms of reference

Purpose of the evaluation

The TOR for the evaluation of the Student Centred Teaching in the South Pacific Project had the following well formulated purpose:

The evaluation is intended to assess the impact of the project. The evaluation should draw conclusions on the key project successes and challenges to date, so that lessons learned may be integrated into future programme development, improvement and expansion.

NZAID envisages the findings of this evaluation will be used by project partners to guide future delivery of this initiative, in particular the current Funding Arrangements (2006-20011). The knowledge and information gained from this evaluation will be fed back into relevant policies, strategies and activities of key partners in order to ensure on-going improvement to programme delivery, future development of impact, efforts towards sustainability and alignment with relevant sector strategies.

NZAID will use the findings to guide decision making around future inputs into regional educational initiatives in the Pacific.

Reporting requirements and feedback to stakeholders

The TOR for the end of project evaluation of the Pacific Health Research Council clearly listed the reporting 'outputs' required, and then presented a table of key dates, events, milestones and fees. While these reporting requirement were light on some details, and did not describe what would happen to the final report, or what should be included in the report, they described succinctly what was expected in terms of outputs (including reports, workshops etc). Furthermore it was clear how feedback to stakeholders would occur. The reporting requirements were listed as outputs:

Output 1: *A verbal report of preliminary findings to the Evaluation Advisory Group (EAG)*

Output 2: *A consultation workshop to report back evaluation findings to key stakeholders.*

Output 3: *A draft written report to the EAG within ten working days from completion of field work.*

Output 4: A final written report, submitted within ten days of receipt of comment and feedback from the EAG.

DAC criteria addressed

The TOR for the evaluation of the RED Cross International Delegates Programme was on of the very few TOR that referred to The DAC Quality standards and/or NZAID documents. The TOR stated:

NZAID will make available to the evaluator ‘OECD DAC quality standards for evaluation’ as well as NZAID’s ‘Evaluation Policy Statement’ and ‘Guideline on the Structure of Evaluations’ in order to inform the methodology and reporting of the evaluator.

2. Evaluation and Review Reports

Methodology

Design of methodology (1)

The review of the ‘Building an Infrastructure for Research in the Cook Islands and Pacific’ initiative provided an excellent example of appropriate and well designed methodology.

The methodology stated that it was guided by “Pacific research principles (HRC 200%) and by core values outlined in NZAID documents (these were described in a footnote). It then clearly described the methodology.

An annex was included with the design of the methodology and work plan.

A table listed the evaluation questions as per the TOR, and beside each evaluation question, the methods used to gather information to answer the questions. For example, the first two rows of the table:

<i>Terms of reference</i>	<i>Method</i>
<i>To what extent is the initiative aligned with the priorities of the Cook Islands Government?</i>	<i>Review of documentation: National Research Council: Development Plans Discussions with government agencies including AID management Team and Line ministries</i>
<i>What are the strengths and weaknesses of the project? What is going well? What is going less well and why?</i>	<i>Interviews/focus groups with government, NGO and private sector agencies</i>

Design of methodology (2)

The review of Volunteer Service Abroad (VSA) also included the work plan in an appendix. This work plan clearly described the methodology. It also included the principles on which the review was based:

The review will be consistent with the principles specified by NZAID and will:

- *involve stakeholders at all stages of the process*
- *build capacity as an integral part and intended outcome of stakeholder involvement in the process*

Design of methodology (3)

The methodology described in the Pacific Health Research Council review was a good example of a detailed description of how the review was conducted, leading to findings that were reliable and comprehensive.

The methodology states that the methodology was guided by NZAID core values (which were listed). It included full description of who interviews were conducted with (both those listed in the TOR and additional stakeholders). The questions asked in the interviews and the questionnaire used for stakeholders to complete prior to telephone interviews was also included as appendices. Stakeholders who did not respond to requests for interviews were listed. Reasons for not using questionnaires for quantitative analysis were described.

Limitations of the evaluation/review described

The limitations of the Pacific Health Research Council review were clearly described, this adding to judgement on the reliability of information in the review. For example:

The timeframe for completing the evaluation was very short

Second reviewers were reliant on the list of key stakeholders provided by NZAID. The reviewers also contacted other key individuals as advised by key identified stakeholders'

Third, the reviewers were reliant on information made available by NZAID and secretariat of PHRC, for document analysis. This may have resulted in some important documentation being omitted from the review'.

Fourth, reviewers could not contact any health minister for interviews

Finally the TOR required reviewers to visit three PICs and the PHRC has MoUs with 11 countries. This could have resulted in some information being missed from the review.

Using direct quotes to provide depth and understanding to qualitative data

The review of the Pacific Health Research Council was one of the only evaluations/reviews that used direct quotes from stakeholders to illustrate findings. The quotes gave the reader an idea of the tone of the interviews and provided 'depth' to the findings. For example the review stated:

The ongoing conflict that has existed between members of the PHRC Executive Committee and FSM, has had the consequential effect of impacting on the delivery of outputs: 'It used up all our energy', 'we have become dysfunctional together as partners and too much of our meeting times has been spent on this rather than the work needed'.

EC members expressed the opinion that 'we need to start afresh ... a new start (2)...'

Independence of evaluator described

In the Review of the Himalayan Trust the reviewer clearly stated their independence. This was one of the only evaluations and reviews that addressed this issue:

I was accompanied by Ang Rita Sherpa, the chief administrative Officer of the Trust in Nepal, and he acted as an interpreter on the occasions when this was necessary. Ang Rita was responsible for setting the itinerary for my visit, but I suggested several of the interviews myself and had a number of meetings on my own, so I am confident in the objectivity of the findings of the review.

Participatory evaluation

The monitoring review of the Philippines National Ecotourism programme was an example of a review that took a participatory approach. The participatory approach was described in the TOR which were included in the report (not in the report itself).

The participatory approach included working collaboratively with the MSC and in-country counterparts in the review, considerable consultation over the processes of the review, input to the review from a range of stakeholders, including tourism operators, community-based organisations and local community (men, women and youth), and feedback on the draft report from Government of Philippines (as well as NZAID and the MSC) to be included in the final report.

The report itself clearly outlined the stakeholders that had engaged in the monitoring review. The method used to find information were stated as interviews and workshops (although further details on these were not included).

The outcome of this participatory approach was that local stakeholders apparently felt 'the review recommendations were good and a solid basis for the project to move forward. They also appreciated the participatory approach under which the review was undertaken.... The participatory process has resulted in strong local uptake of review recommendations' (ERC submission).