

AusAID

Study of
Independent
Completion Reports
and other evaluation documents

Commissioned in support of the
Independent Review of Aid Effectiveness

Study undertaken by Peter Bazeley

With research assistance and data analysis provided by
Phil Cowling, Emma Haegeman and Craig Mathieson

Final Report | March 2011

The views contained in the paper are those of the author and are
not indicative of the views of the Australian Government or the
Review Panel.

Study of Independent Completion Reports and other evaluation documents

Commissioned in support of the Independent Review of Aid Effectiveness

Study undertaken by Peter Bazeley¹ with research assistance and data analysis provided by Phil Cowling², Emma Haegeman² and Craig Mathieson²

Final Report | March 2011

Contents

| | |
|--|-----|
| Abbreviations & Acronyms | ii |
| Summary | iii |
| Part 1: Introduction | 1 |
| Part 2: Methodology | 2 |
| Context | 2 |
| Methodology applied | 3 |
| Limitations | 4 |
| Part 3: Findings | 5 |
| A. Scope and quality of evaluation | 5 |
| B. The quality of activities evaluated | 9 |
| Primary rating data | 9 |
| Potential contributors to quality | 12 |
| More qualitative analysis | 16 |
| Part 4: Discussion and conclusions | 19 |
| Program performance | 19 |
| Process of evaluation | 21 |
| Acknowledgements | 23 |
| Annex 1: Terms of Reference | 25 |
| Annex 2: Template for ICRs..... | 27 |
| Annex 3: Hypothesised contributors to aid effectiveness..... | 30 |
| Annex 4: Summarised illustrative activities..... | 32 |

¹ Peter Bazeley Development Consulting, Dorset, UK | e-mail: peter.b@zeley.com

² theIDLgroup, Bristol, UK | e-mail: info@theIDLgroup.com

Abbreviations & Acronyms

| | |
|----------|--|
| AidWorks | AusAID's central management information database |
| ACR | Activity Completion Report (produced by a managing contractor) |
| ADB | Asian Development Bank |
| AUD | Australian dollar |
| AusAID | Australian Agency for International Development |
| DAC | Development Assistance Committee of the OECD |
| GPF | Global Partnership Fund |
| ICR | Independent Completion Report |
| IPR | Independent Progress Report |
| M&E | Monitoring and Evaluation |
| MDGs | Millennium Development Goals |
| NGO | Non-Government Organisation |
| ODE | AusAID's Office of Development Effectiveness |
| OECD | Organisation for Economic Cooperation and Development |
| PFM | Public Financial Management |
| PICTs | Pacific Island Countries & Territories |
| PNG | Papua New Guinea |
| 'Post' | In-country AusAID office / representation |

Summary

The purpose of the study is inform Australia's Independent Review of Aid Effectiveness by examining what AusAID evaluations tell us about efficiency and effectiveness, and by assessing the quality and utility of the evaluation process itself.

We examined 162 evaluation reports – principally *Independent Completion Reports* (ICRs) and *Independent Progress Reports* (IPRs) – on AusAID activities completed since 2006/07, representing just over 2/3^{ds} of registered evaluations over that period.

Quantitative assessments were made utilising quality ratings awarded at evaluation (and proxy ratings awarded under this study where these were not provided) and our own scoring of what reports told us of the significance of hypothesised key 'contributors' to aid effectiveness. The apparent significance of those 'contributors' in the activities were then cross-correlated with the ratings awarded against the principal DAC and AusAID evaluation criteria.

More qualitative analysis of the performance story that the reports were in aggregate telling was also logged.

Inconsistencies and weaknesses were seen in the quality and utility of evaluation reports, most significantly because of their tendency to focus on activity and process rather than outcomes and impacts – perhaps because ICRs are principally configured to validate implementing agents' own activity-level reports. The interpretation of DAC and AusAID evaluative criteria was also often less than demanding. Many reports noted that the quality of monitoring and evaluation data compromised the scope of the evaluation, particularly as regards examining outcomes and impacts.

Nonetheless the study highlighted important features of, and contributors to, the effectiveness and efficiency of the aid program.

Overall, over three-quarters of evaluated activities (for which there were reports) were rated as being of satisfactory quality or better. 24% and 5%, respectively, were rated as 'good' or 'very high' quality. No activity for which there was a report was rated overall as 'poor' or 'very poor'.

Quality ratings for specific evaluative criteria were more mixed, and suggested diverse causes. Relevance was rated particularly well, but possibly because the interpretation of 'relevance' in AusAID evaluations is narrow and procedural. Sustainability was rated as satisfactory in only two-thirds of activities. Ratings for gender equality, and in particular the quality of monitoring and evaluation, were low.

Our assessment of contributors to quality suggested that well-contextualised design with strong ownership and leadership by partner governments, and intelligent, analytical and responsive implementation (including a strong role for Post in policy dialogue) were the principal – if unsurprising – drivers of effectiveness. There is close correlation between greater use of government systems (loosely defined) and sustainability.

Foggy objectives and a preoccupation with activity and process (as opposed outcomes and impacts) were the principal contributors to suboptimal performance.

The data suggest that AusAID is yet to make the most of its role and influence in multi-donor (including multilateral) partnerships.

A number of comments are made about the need better to configure and position monitoring and evaluation as a tool for assessing effectiveness and efficiency at outcome level, and in making strategic programming choices for the purposes of increasing aid effectiveness.

Study of

Independent Completion Reports and other evaluation documents

Commissioned in support of the Independent Review of Aid Effectiveness

Study undertaken by Peter Bazeley³ with research assistance and data analysis provided by Phil Cowling⁴, Emma Haegeman⁴ and Craig Mathieson⁴

Final Report | March 2011

Part 1: Introduction

1. An Independent Review of Aid Effectiveness⁵, the first for fifteen years, is examining the effectiveness and efficiency of Australia's aid and will guide the future direction of the aid program. **This study of AusAID Independent Completion Reports** (Terms of Reference at Annex 1) was commissioned to inform the Independent Review by seeing what we can learn about the effectiveness of aid activities from recent-past⁶ AusAID activity-completion evaluations, and by assessing the quality and utility of that evaluation process itself.

2. Independent Completion Reports (ICRs) are the product of, typically, two to three week evaluations at the completion of AusAID-funded activity, undertaken by external, AusAID-contracted, evaluators. With some exceptions, ICRs should be produced for all activities exceeding a financial threshold of AUD 3 million. AusAID requires ICRs to report against the standard OECD Development Assistance Committee (DAC) evaluation criteria⁷ of relevance, effectiveness, efficiency, impact and sustainability as well as three AusAID-specific criteria: gender equality, monitoring & evaluation and analysis & learning. AusAID's template for an ICR, which includes prompts as to how to interpret these evaluation criteria, is reproduced at Annex 2.

3. From April 2009 ICRs started to include a standardised 1-6 quality rating against the principal evaluation criteria (with the exception of 'impact'): these ratings formed much of the quantitative basis of this study. ('Methodology', below.)

4. Protocols have also been established to monitor the quality of evaluation reports. Some but not all recent evaluation reports have been subject to a double-blind

³ Peter Bazeley Development Consulting, Dorset, UK | e-mail: peter.b@zeley.com

⁴ theIDLgroup, Bristol, UK | e-mail: info@theIDLgroup.com

⁵ See <http://www.aidreview.gov.au/termsreference/index.html>.

⁶ The study looked at reports for financial years 2006/07 to 2009/10 inclusive.

⁷ See http://www.oecd.org/documentprint/0,3455,en_2649_34435_2086550_1_1_1_1,00.html.

'Technical Review' and awarded a 1-5 rating⁸ for their overall quality. That rating also determines whether or not an evaluation report will be deemed by AusAID to be publishable.

5. For this study, AusAID identified for review 108 ICRs undertaken between July 2006 and June 2010, although reports for 31 of those could not be traced.

6. Other reviews and evaluations were also made available to the study, including Independent Progress Reports (IPRs) (conducted at the mid-point of implementation for shorter activities or at intervals during implementation of longer activities), cluster and thematic evaluations, ODE case studies and other ad hoc reviews and evaluations. Again, not all the reports for listed studies could be traced.

7. In all, the potential and actually-available number of reports comprised:

| AusAID activity reviews and evaluations July 2006 to June 2010 | | |
|---|-------------------------|------------------------------|
| Type of report | Listed by AusAID | Provided to the study |
| Independent Completion Reports (ICR) | 108 | 77 |
| Independent Progress Reports (IPR) | 62 | 41 |
| Cluster evaluations | 7 | 7 |
| Joint evaluations | 5 | 5 |
| ODE case studies, etc. | 6 | 2 |
| 'Impact studies' | 2 | 2 |
| Thematic evaluations | 4 | 1 |
| 'Other' / unclassified evaluations | 42 | 27 |
| TOTAL | 236 | 162 (69%) |

Part 2: Methodology

Context

8. The methodology adopted for the study had to relate to a number of givens:
- i. This had to be a rapid review: the study was to inform a wider review process with a short timeframe.
 - ii. While the sample size (initially 236) was comparatively small compared to the number and diversity of activities completing during the same period (nearly two thousand), it was still large in terms of sheer volume of text: the intended number of reports to be reviewed and the time available suggested we could allocate no more than 45 minutes to each report.
 - iii. Reports were known, *ex-ante*, to be diverse and variable in their scope, content, style and quality.

⁸ Score 1 = 'Report requires complete revision'; 5 = 'Report is of excellent standard'. Only reports with a rating of 3 or above are deemed publishable.

- iv. Quantitative evaluation ratings were only introduced in the latter part of the period covered by the study and are known to be being applied with some inconsistency.
- v. The protocol for the regular 'evaluation of evaluations' (Technical Reviews), which might have provided a degree confidence to evaluations and quality ratings, was also known to be yielding variable information – to the extent that it has been placed on hold pending its own review.

Methodology applied

9. Ratings awarded in ICRs and other evaluation reports against the DAC criteria of **relevance, effectiveness, efficiency and sustainability** and the AusAID criteria of **gender equality, monitoring & evaluation, analysis & learning**, and '**overall quality**', formed our basic unit of our analysis. Where such ratings had *not* been awarded by evaluators we awarded **proxy ratings**, where relevant, based on our reading of what the report seemed to be saying about each of those criteria. AusAID's established evaluation rating scale implies the following assessments:

| Standard AusAID Evaluation Ratings | | | |
|------------------------------------|-------------------|------------------------|----------------------------|
| Satisfactory | | Less than satisfactory | |
| 6 | Very high quality | 3 | Less than adequate quality |
| 5 | Good quality | 2 | Poor quality |
| 4 | Adequate quality | 1 | Very poor quality |

10. In addition, we ourselves (for this study) applied a separate scoring system (para 41 below) to the significance, in the activity being reported, of what we hypothetically considered to be important underlying **contributors** to relevance, effectiveness, efficiency and sustainability. The selected hypothetical contributors are set out in Annex 3 and included such measures as (for example) the extent to which an activity was based on a solid theory of change (relevance), was intelligently implemented and represented a 'learning organisation' (effectiveness), balanced financial and development risks (efficiency) or was essentially transactional or transformational (sustainability).

11. We also screened reports for:

- The time allocated to the evaluation;
- Whether the timing or methodology of the evaluation was reported as an issue;
- The extent to which value-for-money was assessed;
- Whether or not the activity's monitoring and evaluation system was reported to have provided the evaluators with an appropriate quantity and quality of data and information for the evaluation.

12. A two-stage analysis was then undertaken of the 162 available reports:

Firstly to:

- Screen the reports for their fitness for purpose (some were discarded as being too unique or too incomplete to be useful);
- Award proxy ratings, where needed/possible, against the DAC and AusAID evaluative criteria (para 9 above);
- Assess and rank the 'contributors' (para 10 above);
- Log scores onto a database;
- Log interesting and emerging issues onto a rolling blog shared among the study team.

Secondly to:

- Cross-correlate quantitative data for trends and patterns and any obvious cause-and-effect relationships; and
- Draw qualitative conclusions about the aid effectiveness story that the evaluations were painting.

Limitations

13. Data limitations emerged in terms of:

- The diversity of activities, in every combination of every dimension: sector, objectives, modality, geography, management and implementation arrangements, partnerships, scale and duration, and approach to evaluation. 'Comparing apples with oranges';
 - Often leading to a very low sample size for any one set of activities or evaluations.
- Variable interpretation among evaluators of DAC criteria.
- Incomplete ratings.
- A narrow normal distribution of ratings, congregated around the mid- ('just satisfactory') range. The sample size for outlying observations was mostly low.
- Disparities, seemingly, between the narrative of a report and the ratings that were subsequently awarded: evaluators would often sometimes some aspect of the activity that was clearly less than satisfactory, but then award a neutral or better rating. (See text box below: 'How good is good enough?')
- And, as discussed later, an overwhelming tendency for reports to be telling a story about low-level processes rather than higher-level outcomes and impacts.

14. Note that where the sample size was particularly small, rounding discrepancies will be seen in the percentages quoted in some of the charts that follow.

Part 3: Findings

15. While our principal interest is in what the evaluations tell us about the effectiveness and efficiency of AusAID activities, we present first data on the coverage and quality of evaluations, as this by implication provides some qualification for the later discussion on the quality of the activities.

How good is good enough?

The *Independent Review* is intended to “make a strong aid program even better”, so we have tended to interpret the benchmark of quality as lying at the upper end of the ratings applied to evaluations and evaluation reports.

A. Scope and quality of evaluation

16. Of the reports provided to the study, about half contained quality ratings against some or all DAC/AusAID evaluation criteria (with some inconsistency, especially for gender). Note that ‘impact’ is not rated in AusAID ICRs and IPRs. There were also quality ratings registered on AidWorks for a further 30 activities, but for which the original reports could not be traced. We included these ratings in our analysis of data where applicable.

17. The ideal, ‘gold standard’, evaluation for this study would be one where:

- A standard-format independent completion evaluation had been undertaken; *and*
- Where a report was available; *and*
- Which contained quantitative ratings for the quality of the activity; *and*
- Where the report itself had also been subject to ‘Technical Review’ (peer review), *and*
- Where the evaluation report was deemed to be of acceptable quality. (Thus leading to the highest level of confidence in the data).

18. There were 35 such ‘gold standard’ reports, some of which feature among the example activities summarised at Annex 4.

How reliable and representative is the sample?

19. Of the 1,999 activities funded by AusAID from July 2006 to June 2010 (representing an actual expenditure of over AUD 9.4 billion), 547 activities (representing over AUD 8.5 billion) were marked for performance-tracking and should have been subject to an ICR⁹. Of those 547, the Review Secretariat established that 108 had undergone an ICR (equating to about 20% compliance), although almost a third of those reports could not be traced for this study.

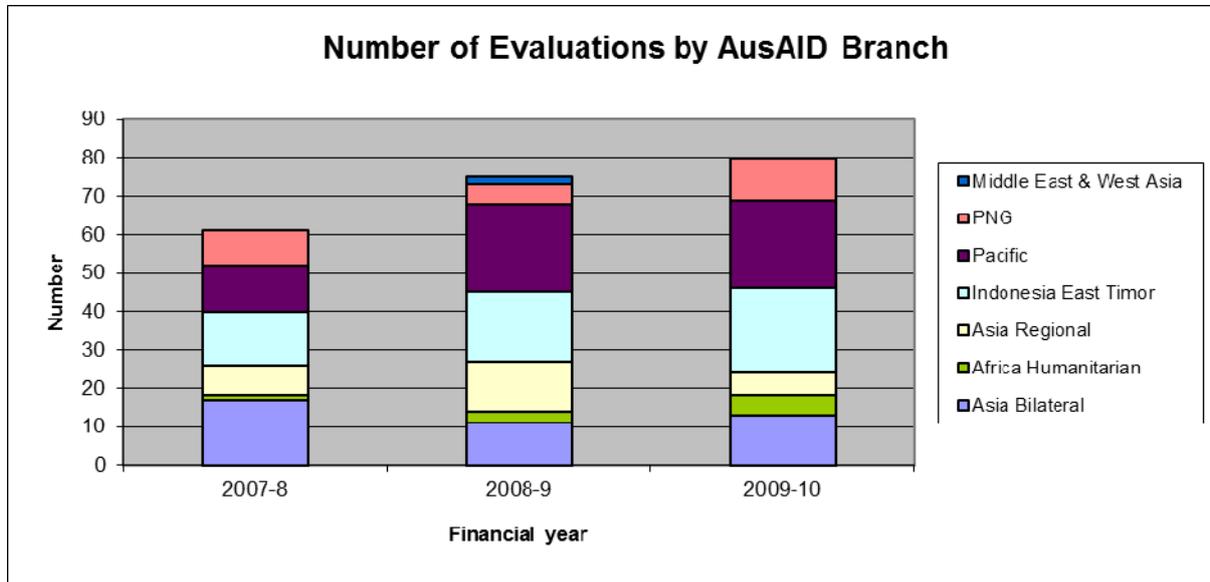
20. Other forms of evaluation have however also been completed (which may have exempted some activities from the requirement for an ICR) and many of these were also made available to the study. (Para 7 above.) In all, the study received 162 reports.

21. Putting aside the availability of a report, the number of evaluations undertaken each year is increasing across most AusAID Branches, but remains below the number

⁹ The great majority of AusAID activities are of small value and fall below the ICR-requirement threshold – see para 2.

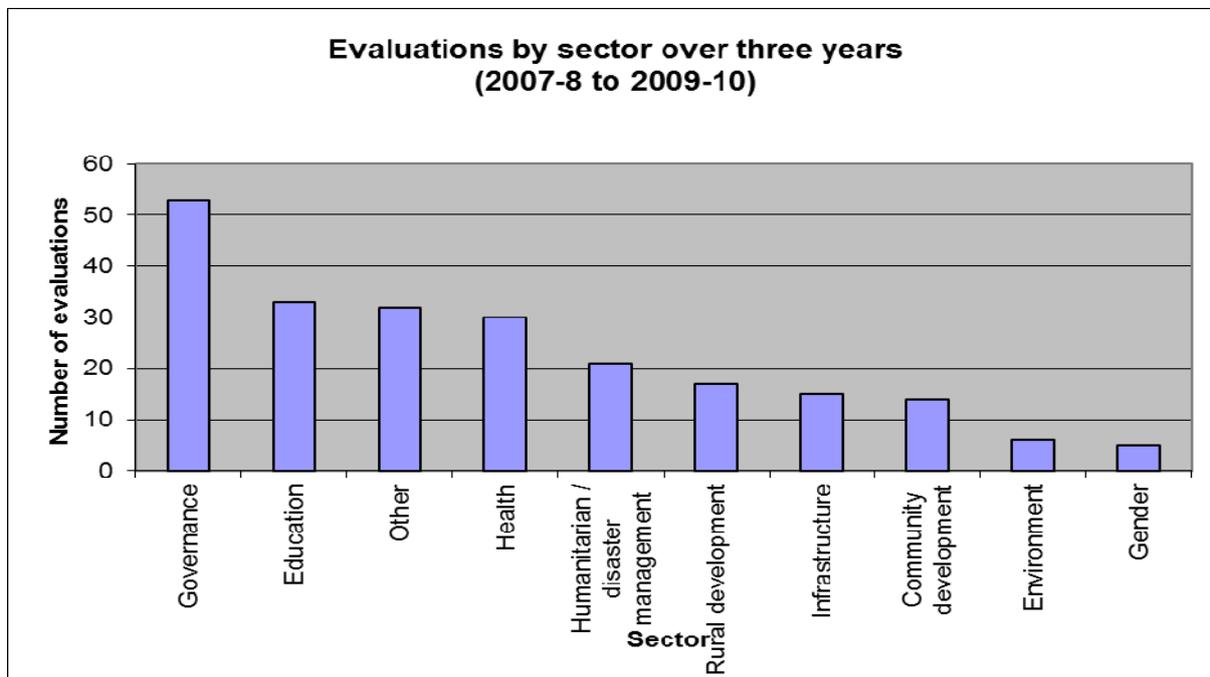
that is required by current AusAID protocols to be evaluated on completion. (Chart 1 below.)

Chart 1



22. Some sectors are represented better than others¹⁰ (Chart 2 below):

Chart 2

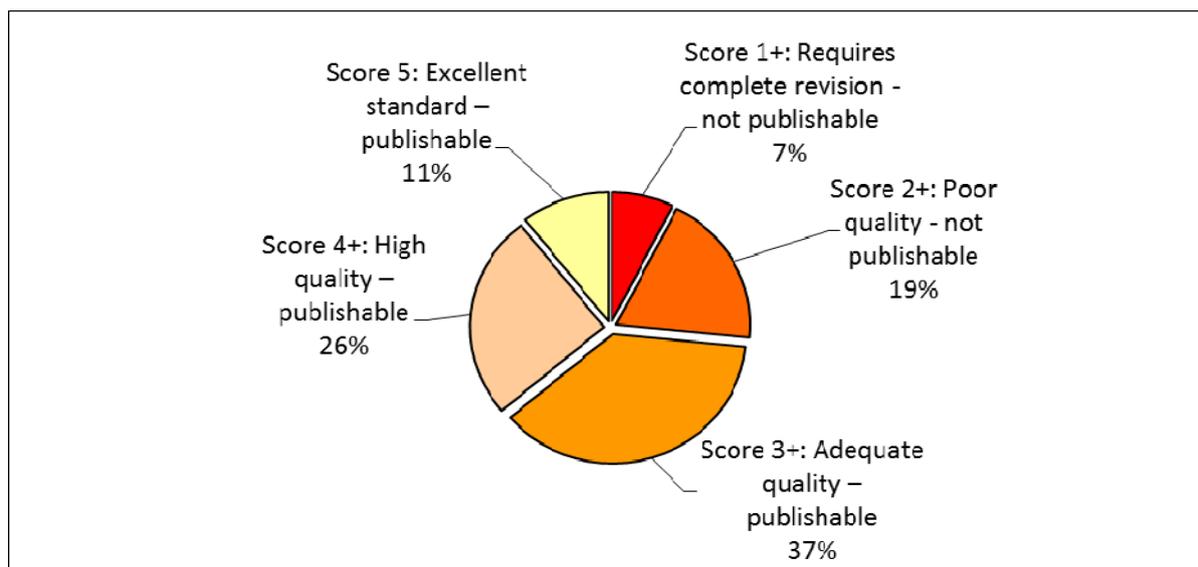


¹⁰ Although this needs to be compared to the numbers of activities funded in each sector, which data we did not have.

How good are the reports?

23. Ninety-eight of the evaluation reports provided to the study (mostly ICRs and IPRs but also some others) had been subject to a Technical Review by peer evaluators and the evaluation report itself also rated for quality. Of those, about three-quarters were deemed to be of adequate quality. But – to put it another way – over a quarter were deemed to be of insufficient quality to be published. Only 11% were assessed by peers as being of an ‘excellent’ standard. (Chart 3 below.)

Chart 3
Quality of Evaluation Reports – ‘Technical Review’ Ratings

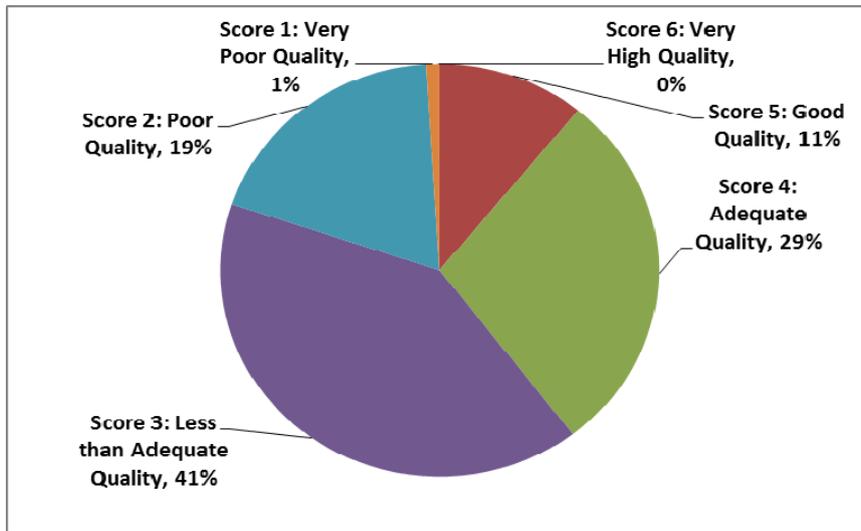


Were the right data there in the first place?

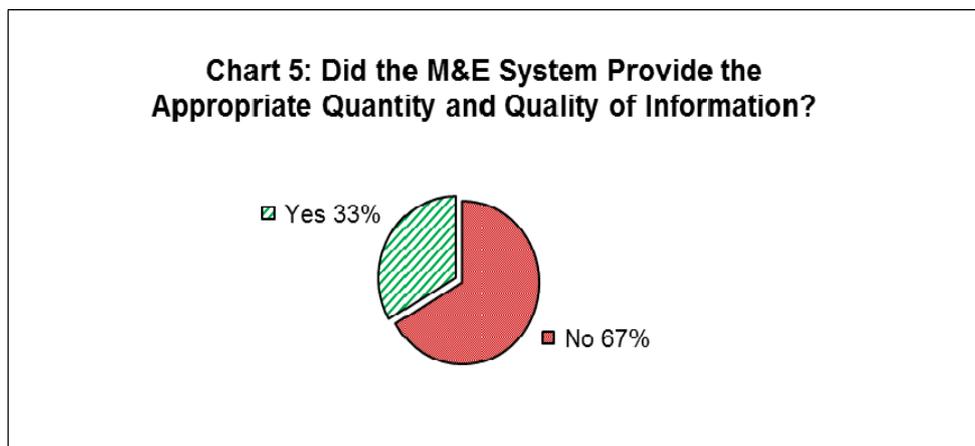
24. ICRs principally validate performance data provided by an activity's implementing agency: they rarely undertake primary data collection themselves. The robustness of ICRs and most other evaluations are therefore highly dependent on the quality of the activity's own baseline data and monitoring and evaluation systems.

25. In terms of the AusAID evaluation criterion of quality of 'Monitoring & Evaluation' (para 9 above), over 60% of ratings were 'less than satisfactory', including 20% deemed 'poor' or 'very poor'. No activity's monitoring and evaluation was rated 'very high quality'. (Chart 4 below.)

Chart 4
Ratings Awarded in Evaluation Reports to the Quality of Activities' Monitoring & Evaluation



26. In our own assessment of the contributors to effectiveness (para 10 above), we noted that of the 58 reports that discussed the quality of monitoring & evaluation (M&E) **data**, two-thirds (Chart 5 below) indicated that they were insufficient properly to fulfil the needs of evaluation – particularly at the level of analysing outcomes and impacts.

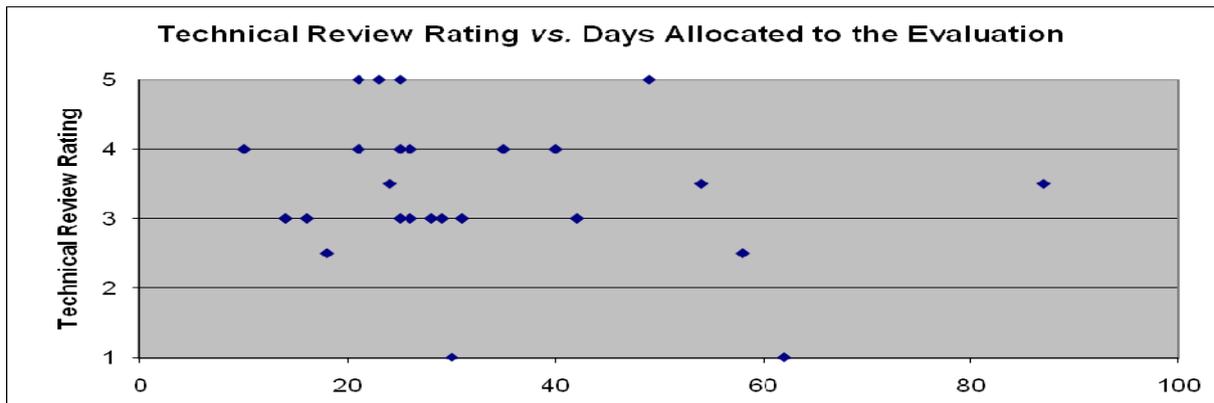


Time allocation

27. Reports often made reference to the time allocated to the evaluation as influencing, to some extent, its scope and quality. However there was no obvious correlation in the bulk of the evaluations between the time allocated to the evaluation (in those evaluations) and subsequent quality-of-evaluation ratings¹¹. The reports of some large meta-evaluations with a long timeframe exhibited slightly lower average Technical Review ratings. (Chart 6 below.) Excluding these long meta-evaluations, the average time allocated to evaluations was 23 days.

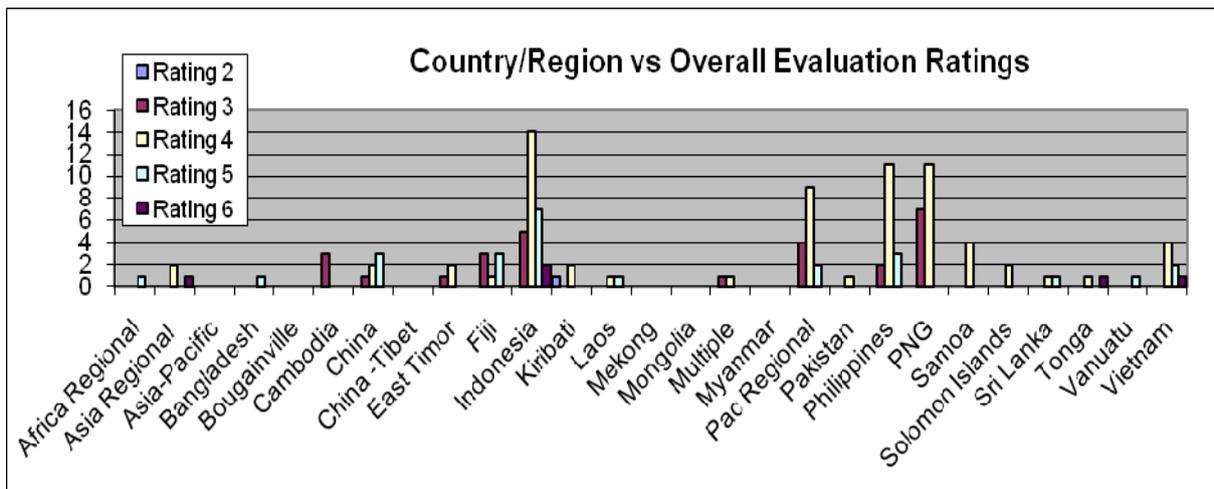
¹¹ The data for this cross-correlation were questionable: it was not always clear whether figures for the time allocated to evaluation were calendar time or team person-days.

Chart 6



28. Geographic coverage and the number of evaluations completed appears to reflect, in the main, the relative size of AusAID bilateral, but perhaps not regional, programs (Chart 7 below):

Chart 7



B. The quality of activities evaluated

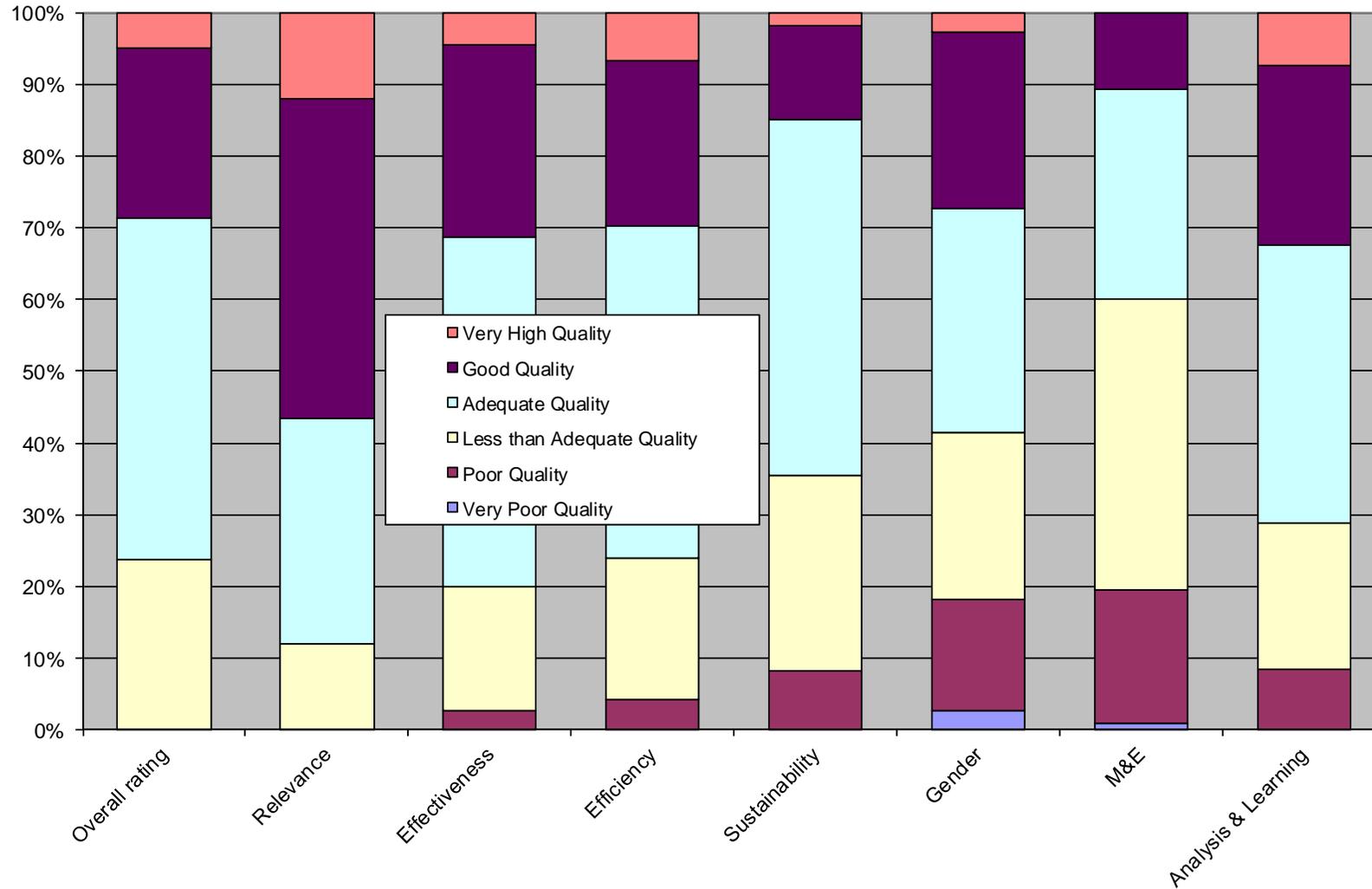
29. With the caveat that, as discussed in Part A, the data may not be as robust and comprehensive as first meets the eye, we now turn to examining how the quality of AusAID activities was evaluated in the ICRs and other reports.

Primary rating data

30. Chart 8 overleaf displays the distribution of the primary AusAID ratings applied to four of the five¹² DAC and four AusAID criteria now required to be awarded in AusAID evaluations.

¹² The DAC criterion of 'impact' is not rated quantitatively in AusAID ICRs, but is sometimes discussed in their narratives.

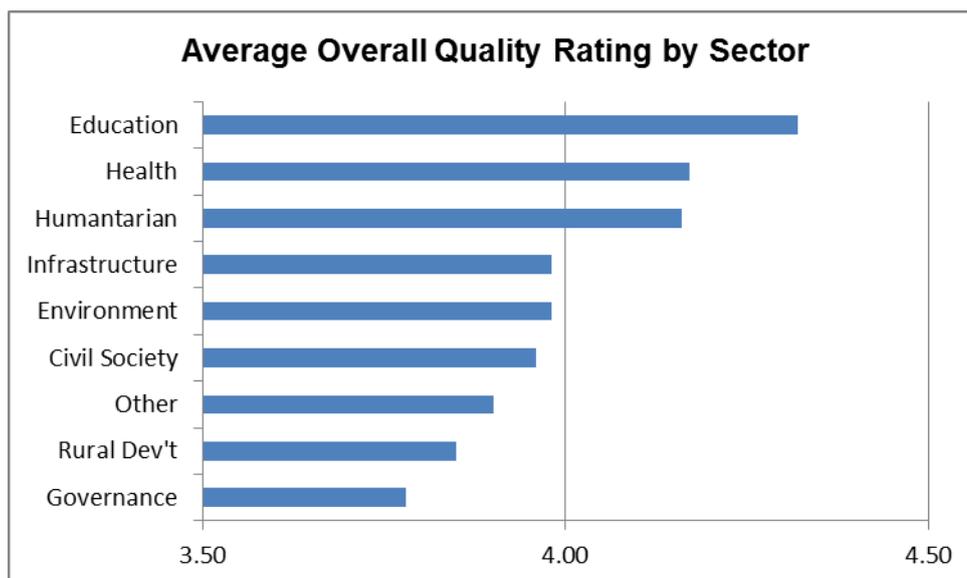
Chart 8: Primary AusAID Ratings by Evaluation Criteria



31. **Overall**, over three-quarters of the evaluated activities (for which there were reports) were rated as being of satisfactory quality or better. 24% and 5%, respectively, were rated as 'good' or 'very high' quality. No activity for which there was a report was rated overall as 'poor' or 'very poor'. The most frequent overall rating was '4' ('adequate'), awarded to 48% of activities.

32. There are some marginal differences in the average of the overall ratings¹³ awarded to different sectors, with ratings for education, health and humanitarian activities (which together constitute almost 40% of rated evaluations) averaging more than 4, and other sectors averaging less than 4. (Chart 9 below.)

Chart 9



33. Governance activities constitute almost a quarter of the rated evaluations.

34. Ratings for '**effectiveness**', '**efficiency**' and '**analysis & learning**' follow a similar middle-of-the-road pattern, although all of those start to pick up a small number of 'poor' ratings.

35. More diverse rating patterns are seen under 'relevance', 'sustainability', 'gender' and 'monitoring & evaluation':

36. '**Relevance**' rates well, being deemed to be adequate or better in over 87% of activities and either 'good' or 'very high quality' in the majority of cases (44% and 12% respectively). Relevance of activities was rated less than satisfactory in less than 13% of activities for which there was a report, and in none of those cases was the rating 'poor' or 'very poor'. However, as we discuss later, the meaning of 'relevance' is interpreted variably – and often less than ambitiously – in the reports we studied.

37. '**Sustainability**' is deemed to be 'satisfactory' in two-thirds of activities (or perhaps more alarmingly stated as 'less than satisfactory' in a third of activities), and 'good' or 'very high quality' in only 15% of activities.

¹³ Average overall ratings can be highly influenced by the a few outlying ratings where the sample number is small.

38. **‘Gender equality’** is rated as being less than satisfactory in more than 40% of cases, with nearly half of those being rated ‘poor’ or ‘very poor’. Ratings for gender equality are some of the most diverse.

39. As already highlighted, **‘monitoring and evaluation’** scores badly, with the majority of activities (over 60%) rated as less than satisfactory and with the largest proportion (20%) of ‘poor’ or ‘very poor’ ratings of any category.

Potential contributors to quality

40. While a few headlines emerge from the primary data on outturn evaluation ratings, we also wanted to look beyond those ratings for any consistent underlying *contributors* to quality and aid effectiveness. Later we highlight a number of qualitative observations in this respect, derived from the stories the reports tell, but we first examine some quantitative data relating to the scores we awarded activities in this study in terms of our hypothetical ‘contributors’ to quality. (Para 10 above.) These relate to the principal evaluation criteria of relevance, effectiveness, efficiency and sustainability. (Annex 3).

41. Provided there was some evidence in the report (and except where there was a simple yes/no answer) we applied ‘contribution’ scores as follows:

| Scores awarded by this study against hypothesised ‘contributors’ | |
|--|---|
| Contribution score 1 | Evidence of the contributor playing a significant, positive, part in the performance story, and/or it being particularly well managed |
| Contribution score 2 | Evidence of the contributor playing some positive role |
| Contribution score 3 | Evidence of negative contribution and/or being poorly managed |
| Contribution score 4 | Evidence of inappropriate omission from the activity |

42. We then cross-correlated our ‘contributor scores’ with the primary evaluation ratings (para 9) for the relevant evaluation criteria.

43. In several cases we found no obvious, consistent, correlations between the ‘contributor scores’ we awarded reports and evaluators’ final quality ratings. There may be some merit in undertaking more complex cross-correlations beyond the timeframe of this study. In other cases, our hypothesised ‘contributors’ simply didn’t feature in evaluations’ narratives. However, there do appear to be some correlations in some areas, which are presented below.

44. **In the following series of charts, the horizontal axis represents the primary 1 – 6 evaluation ratings awarded to the activity by evaluators (1 = very poor, 6 = very good). The vertical axis shows the percentage frequency**

Gender

“There has been no gender analysis for the GPF and consequently no consistent monitoring differential impacts on men and women.”

AUD 50m Australia-Indonesia Government Partnership Fund

“There appears to be little awareness among the PICTs (or the Advisers) about the various influences gender can have on project outcomes.
“... there are no project activities designed to ascertain whether men, women and children have different roles”

AUD 12m Pacific Regional Influenza Pandemic Preparedness

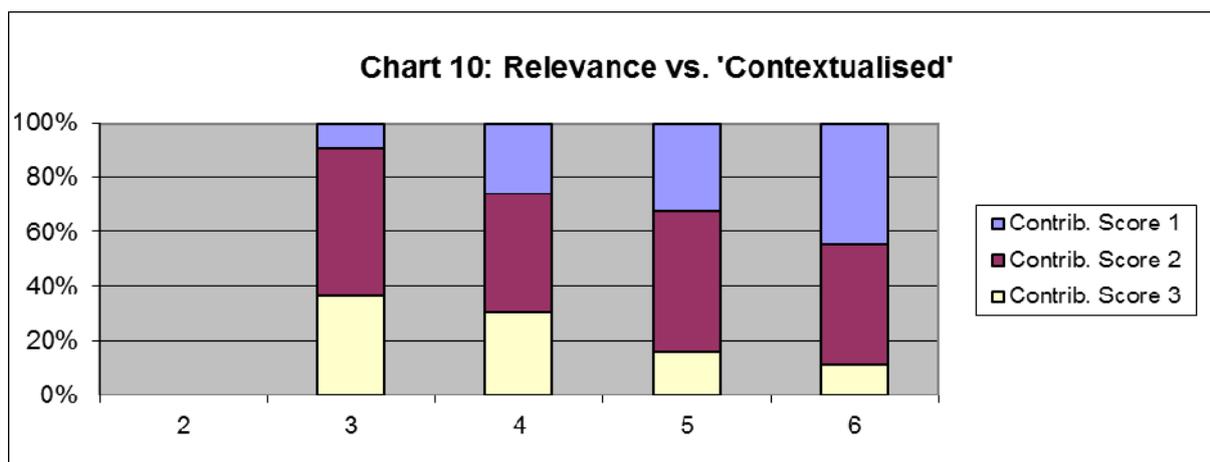
of each of our 1 – 4 ‘contributor scores’ for activities in each of the primary rating bands (para 41 above). More dark blue and red, and less yellow, to the right of the chart suggests strong correlations between the contributor and good evaluation ratings.

Contributors to ‘Relevance’

- How well ‘contextualised’ was the activity?

45. By ‘contextualised’ we mean to what extent did design and implementation explicitly recognise and understand the wider political, social, economic and institutional context, beyond simply saying it was “in line with” stated policies and priorities? To what extent was it assessed as being the ‘right thing in the right place at the right time’ to achieve stated higher-level objectives? Was there a solid ‘theory of change’ that properly recognised the role, strengths and weaknesses of policies, institutions and processes external to the manageable interest of the activity?

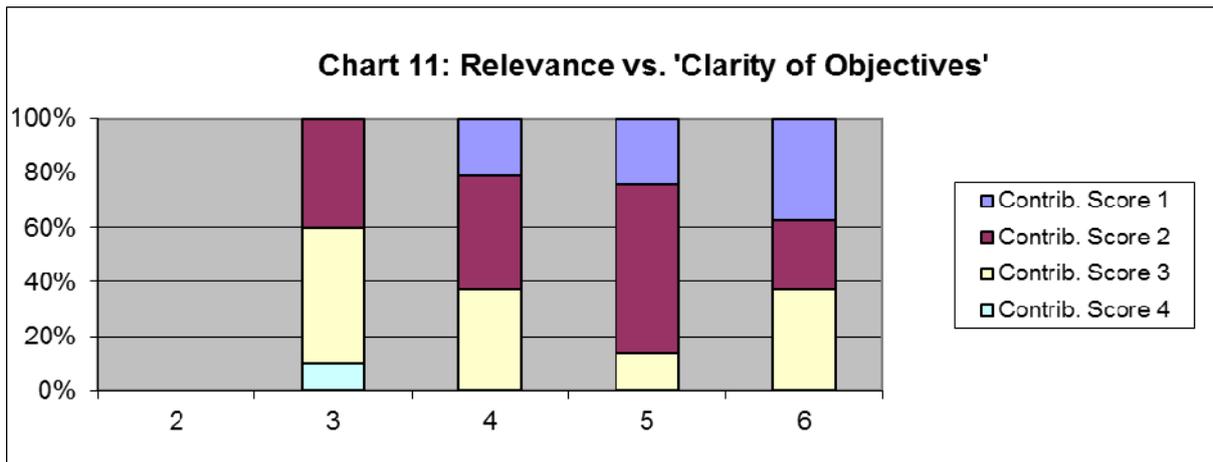
46. As Chart 10 below shows, there appears to be a strong correlation between ‘contextualisation’ and ‘relevance’ with increased frequency of high scores for contextualisation (contribution score 1 and 2) correlating with high primary ratings (5 and 6) for relevance.



- How clear were the objectives?

47. By ‘clarity of objectives’, we mean to what extent were the activity’s objectives clearly set out, reflected a sound intervention logic, were achievable and were measurable. As we discuss later, lack of clarity of objectives and unrealistic objectives feature strongly in evaluation narratives and have a perversely distorting (positive and negative) effect on wider performance ratings: *‘if you don’t know where you’re going, any road will take you there’*. But clarity of objectives is also an important contributor to relevance in terms of ensuring an activity is ‘the right thing in the right place at the right time’.

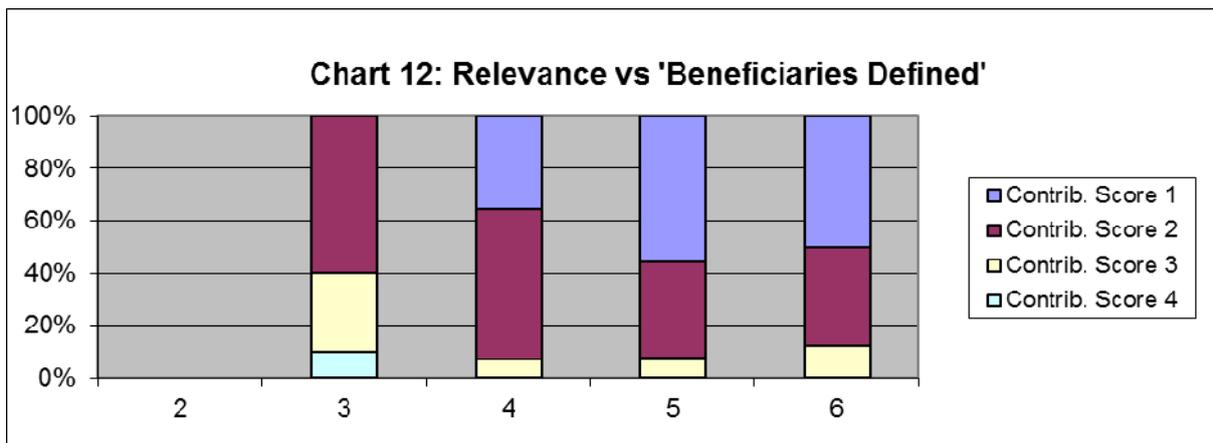
48. Here we again see (Chart 11 below) a strong correlation between evaluators’ commentaries on the clarity of activities’ objectives and final ratings for relevance, but note also that our assessment was that almost 40% of evaluations made adverse comments about the definition of objectives.



- Were the **beneficiaries clearly defined**?

49. By this we mean was there evidence that design and implementation clearly defined who the beneficiaries were to be, and (in particular) the scale of the intended impact? Only then can one say whether an activity is achieving enough, in the right place for the right people, to be relevant.

50. Again we see a correlation (Chart 12 below) between this fundamental element of good design and final ratings for relevance.

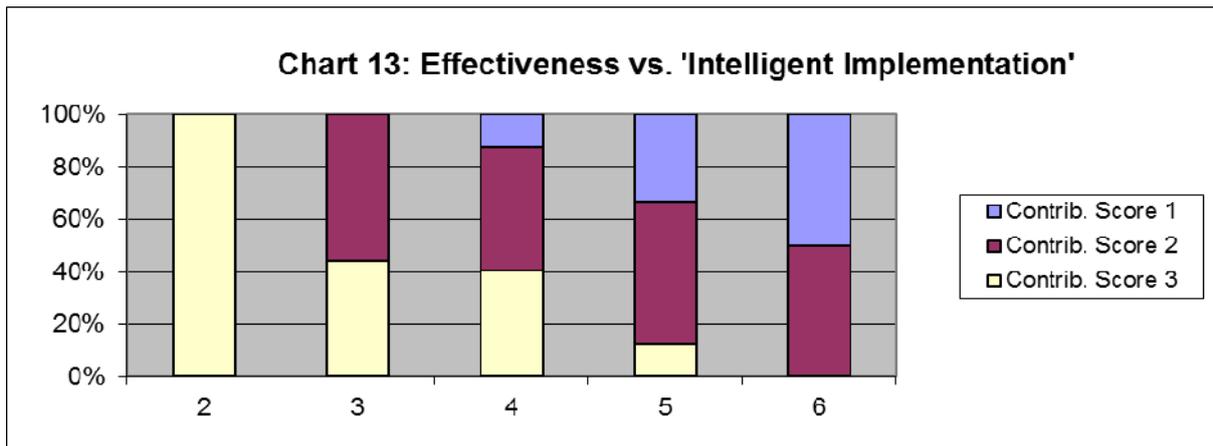


Contributors to 'Effectiveness'

- **"Intelligent implementation"**

51. This curiously-worded contributor to effectiveness is about the extent to which the activity was managed – by both the implementing partner and by AusAID – in a smart, analytical and responsive way, and that there were institutionalised 'sense-making' and feedback loops from monitoring and evaluation back into policy and planning. Did the activity and its management / governance / oversight represent a 'learning organisation'?

52. We see (Chart 13 below) another seemingly clear correlation:

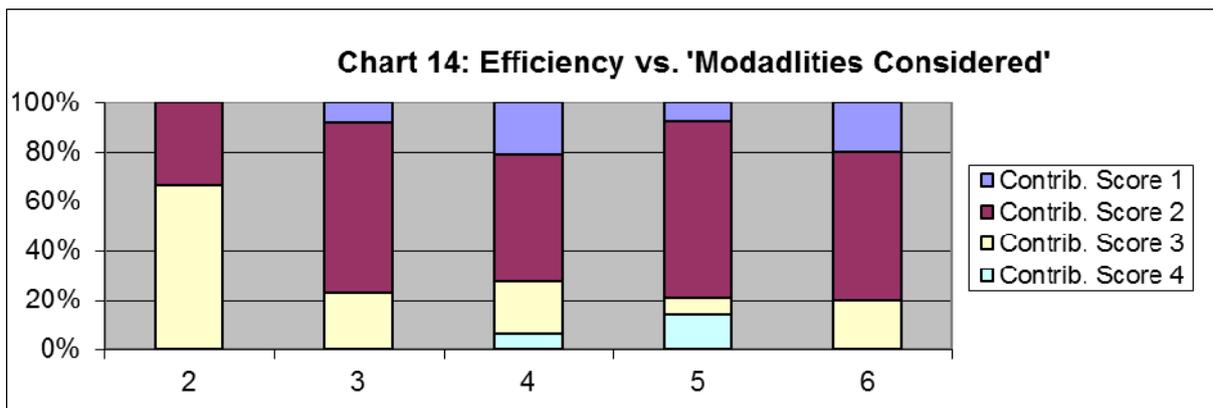


53. Other hypothesised contributors to effectiveness showed more variable correlations that need further analysis. Good scores for gender, lesson-learning and 'coherence' (with existing or earlier initiatives) all exhibited modest correlations with high evaluation ratings for effectiveness, but with some unexplained opposing correlations too.

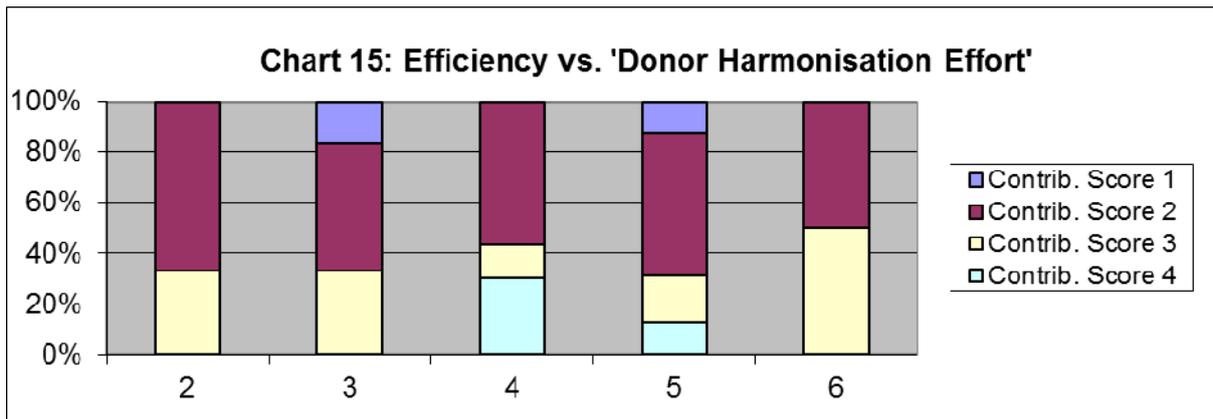
Contributors to 'Efficiency'

54. Correlations with the evaluative criterion of efficiency are not clear. In part, perhaps, because evaluations tended to pitch the analysis of efficiency at low-level activity management processes. We observed that over 30% of evaluations did not, for example, robustly examine value for money.

55. There appears to be a modest degree of correlation, however, between evidence of active discussion of modalities and implementation options in project design (where mentioned in evaluations) and subsequent evaluation of efficiency (Chart 14 below):



56. Interestingly there appears to be some degree of **negative** correlation between donor harmonisation effort and efficiency, which we discuss at para 89 below (Chart 15):

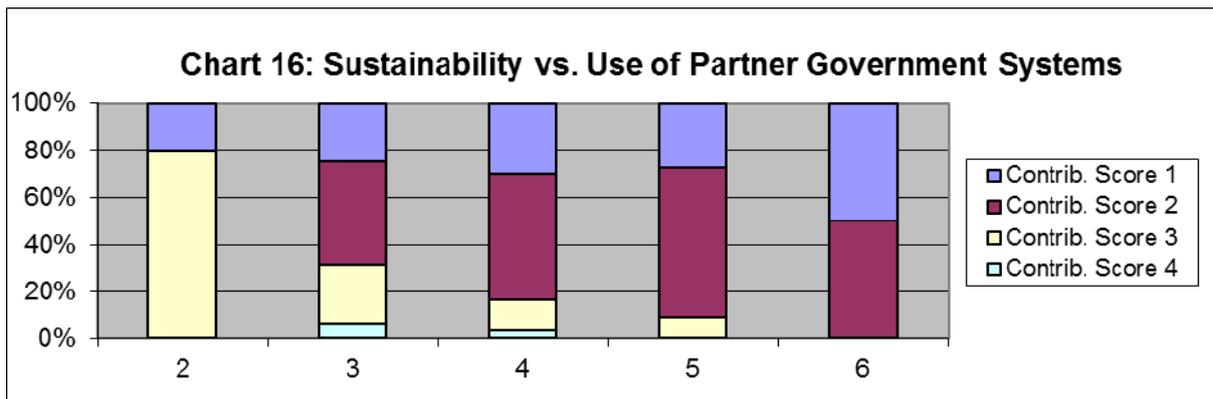


Contributors to 'Sustainability'

57. Para 31 above highlights how the primary evaluation rating for sustainability is less than satisfactory in one third of AusAID-funded activities.

58. One of the central tenets of contemporary international aid effectiveness debates is that sustainability is enhanced by the greater degree of ownership, reduced transactions costs and better alignment with government policies and priorities that comes with greater use of, and integration with, partner government systems. While the dimensions of this are far-reaching and would normally often relate (among other things) to financial flows and partner government financial systems, we interpreted 'use of government systems' in this study loosely to include, also, evidence of working with and through government agencies, or not working through parallel management units, and of partner governments themselves implementing activities.

59. We see (Chart 16 below) a strong correlation between such use of government systems and higher evaluation ratings for sustainability, and a particularly low score for use of government systems among activities whose sustainability was rated by evaluators as 'poor':



Qualitative analysis

60. AusAID has a commendable quality and performance management system in place, extending well beyond the processes and products reviewed in this study. Notwithstanding low levels of compliance, there is much evaluation material to draw on.

61. Reading the evaluation documents reveals few surprises: rather it reinforces much of the contemporary international narrative about what does and does not make

for aid effectiveness. But the evaluations *in aggregate* do provide a qualitative sense of a number of prevalent features, both about the quality of the evaluation process and about the qualities of the activities evaluated.

62. Again we start with the quality of the evaluation process, as this further establishes the limits of confidence we can have in other aspects:

On the quality of the evaluation process

- Disproportionate focus on **low-level** activity and process

63. For reasons we discuss later, the reporting and discussion of input, activity and output dominates evaluations – particularly IPRs and ICRs. They [tend to] tell us whether and how well the activity did what it was supposed to do in terms of processes followed (thus validating implementers' Activity Completion Reports), but not whether it contributed to an effective aid program at the level being discussed now.

- Variable, often somewhat narrow, **interpretation of evaluation criteria**

64. Evaluations exhibit a diverse range of analytical approaches (not necessarily a bad thing) but also interpret the standard evaluative criteria – both the DAC's and AusAID's – variably and more often than not in a somewhat narrow, mechanical fashion. From a 'compliance' and due diligence perspective what evaluators are asked to do is good. But from an aid effectiveness perspective the questions need to be more searching.

- **Institutions** as more than organisations

65. Institutions and institutional reform are often interpreted as being only organisational issues, as opposed to more the complex 'rules of the game' that ultimately determine development outcomes.

- **Too long** and wordy

66. For routine use in this sort of exercise (of informing at an aggregate level the analysis of aid effectiveness, as opposed to activity management and compliance), reports need to be more succinct and more targeted towards contribution to higher-level objectives.

On the quality of the activities evaluated

- **Development professionalism**

67. The most striking qualitative observation, overall, that we would make about what differentiated effective from less effective activities in the evaluation reports studied was the extent to which activities were clearly managed and implemented (and were allowed to be implemented) intelligently (more quantitatively assessed at 51 above) by development professionals exhibiting both **analytical** and sector **expertise**, **cognisant and responsive to the wider political economy** and concepts of **bringing about change** in a **development context**. This applies both to the qualities of the implementing partner (whether that is a managing contractor, a whole-of-government partner, a civil society partner or a unit of the partner government) and the overseeing AusAID management unit.

- **Clarity of objectives**

68. Foggy, loosely-defined objectives are highlighted throughout evaluated activities with, furthermore, a sense that ‘doing good’, as opposed to ‘doing the most strategic thing to effect sustainable change’ is good enough as far as objective-setting goes. (See also ‘transformational’, below.)

- Activities as **learning organisations**

69. Hand-in-hand with these attributes is the observed (or not) ability of an activity to learn and adapt as it is implemented. And to this end, it is the ability of the implementing agent and the overseeing AusAID management unit to be not just responsive but proactive in this respect that shows through as vital. The most effective activities were clearly ‘learning organisations’.

- **Ownership and leadership**

70. Again a striking feature of most the activities evaluated as highly effective was that that there was **genuine and demonstrable ownership** of the intervention by the counterpart organisation or primary stakeholder group. (Beyond those immediately benefitting from project inputs.) Conversely, low levels of ownership were reported as a feature of many of the less effective activities.

71. Strong **leadership** by the partner organisation was repeatedly reported as a feature of the most effective activities.

- Pre-existing **determination to change**

72. The most effective activities built on and supported existing drivers of change (although it was rarely described in those terms), whether those be in the form of individual champions and leaders, progressive organisations, or policy and institutional reforms being decisively driven by their owners.

- **Transformational** vs. transactional

73. By ‘transformational’ we mean that the activity was primarily about bringing about change (para above) to policies, institutions and governance processes such that impacts would be experienced across a population at some scale, above and beyond the direct influence of the aid-funded activity itself. (By contrast transactional activities are those where what is funded itself represents the sum of the benefits – for example the procurement of inputs or the direct funding of services¹⁴.)

74. The most aid-effective and sustainable activities tended to be predominately *transformational* rather than transactional, or (crucially) added value by supporting the transformational components of wider, often loan-funded, investments. While constituting only a small sample size, all the activities rated overall in evaluations as ‘excellent’ were categorised as ‘transformational’ and none as ‘transactional’. However, the more transformational activities also exhibited a wider and more diverse range of quality ratings, suggesting greater development risk (but potentially greater gain).

Clarity of Objectives

“The objective hierarchy and internal logic lacks simplicity, clarity and consistent terminology. The goals and objectives are ambitious, vague and difficult to evaluate. There are few clear, measurable and attributable indicators of success. There is no results framework ...”

AUD 45m ASEAN-Australia Development Cooperation Program

¹⁴ Wider budget support measures, due to their role, function and scale, would need to be assessed differently and were not a feature of this study.

- **Strategies for replication** and roll-out

75. Allied to the issues of 'transformative *vs.* transactional', and also of 'genuine ownership', is the extent to which more transactional initiatives have thought through, agreed and designed-in doable, affordable, demand-led strategies for scaling-up or rolling-out the improvements (typically to service delivery models) brought about by the activity. Evaluations often commented on this negatively, with repeated worries about required partner-government budgetary contributions not materialising.

- **Role of Post** in sector analysis and policy engagement

76. Explicit in some of the most penetrative evaluations is discussion of the role of the AusAID Post in sector analysis and policy engagement, working synergistically with the activity and its implementers. Often good in the best activities and weak in less impactful activities.

- Effective **partnerships with other lead agencies**

77. A number of evaluations comment on less-than-satisfactory, or missed opportunities in, partnerships with the lead agency in wider or bigger programs – often the multilateral loan agencies. A common observation in such discussions is one of AusAID regretting its lack of influence at project design.

Part 4: Discussion and conclusions

Program performance

78. Overall, over three-quarters of evaluated activities were rated as being of satisfactory quality or better. Some of the 'best of the best' activities represent world-class, innovative and probably effective development interventions, demonstrating best practice in aid management. AusAID has the systems, resources and responsiveness to mount quality programs: there is good reason to believe that the objective underlying the Independent Aid Review of 'making a strong aid program even better' is achievable.

79. That said, this study has not been able to draw many conclusions in terms of the efficiency and effectiveness of AusAID activities in contributing to higher-level development outcomes: as discussed below, the principal evaluation tool (the Independent Completion Report) rarely provides such information.

80. There are no big surprises about efficiency and effectiveness at the activity level: what the study demonstrates is that what makes programs work is clarity of objectives, the diligent, informed, application of the established norms of best practice in the design and configuration of aid flows, expert and intelligent analysis of context, responsiveness, and – not least – due attention to the creation of genuine ownership and leadership of development by partner governments. (Not least through greater use of government systems, albeit defined loosely in this study.) It is when those basic principles of good practice slip that programs don't perform.

81. Crucial, this study suggests, is the ability of contractors and Post to engage meaningfully and effectively in, and add value to, partners' processes of making policy and public expenditure choices, and backing those with resources.

82. In short, it's about development professionalism and good donorship creating the space and institutional incentives for best-practice to flourish beyond the exceptional and become the norm – for there are also many examples of where the basic ingredients of good practice are falling short.

Specifically

83. Critically there needs to be greater **clarity of objectives** in the activities that constitute wider programs, related to a solid understanding of context and the external drivers of change. AusAID's new Country Strategy Architecture, with a politically-savvy Country Situation Analysis defining the 'why' of the aid program, can promote this if it is done properly and in sequence.

84. Those objectives need to be clear about **what success looks like** in hard, measurable terms. Evaluations frequently concede that while an activity's objectives were woolly and loosely-defined ("provide support to the strengthening of ...") they can be easily achieved (and therefore have to be scored highly under the ICR template's definition of 'effectiveness'), yet mean nothing in terms of aid effectiveness.

85. Similarly the conceptualisation of '**relevance**' needs to be much smarter, moving beyond being satisfied that (and awarding high evaluation ratings because) an activity is 'in line with' partner government and AusAID policies, to demonstrating – from an aid effectiveness perspective – that the activity represents the right choice of intervention, modality and approach to contribute maximally (among all the possibilities for useful intervention) to the achievement of higher-level development objectives.

86. This more demanding, critical, analysis of what performance and effectiveness means and looks like, with **a more nuanced approach to alignment and engagement with partners'** policy and public expenditure choices, has implications for how Posts work and engage with government and other partners, and for the roles and skills of Post and implementing agents. Invariably, the 'best of the best' activities exemplified such capacity and engagement, and conversely it was often not so apparent in the less impactful activities.

87. A common feature of evaluations is an observation that partner government funding and budgetary allocations fall short of expectations. But this needs to be seen as symptom of the quality of engagement and alignment, not a complaint. If the ultimate proxy test of partner governments' priorities is budget allocation (which it may reasonably be), then the issue may be one of AusAID not having established what those priorities really are, and not having agreed with governments what success will look like.

88. While **capacity-building** features heavily in many activities evaluated, it is rare to see this set in a genuinely institutional sense: more as a head-count of individuals benefitting. But capacity building is an institutional thing: it is about the ability of organisations to fulfil mandates and respond to identified – often shifting – policy and

Capacity Building

“... failed to develop a clear strategy on what capacity the education system needed in the short, medium and long term in order to achieve [national] goals and the MDGs”
AUD 128m Education Capacity Building Program, PNG

institutional requirements and will comprise many different facets beyond, simply, staff skills.

89. The study suggests, but somewhat inconclusively, that AusAID is **better at doing its own thing** than working with other donors and multilaterals, and that it is sometimes disappointed, in retrospect, with its influence on the quality of wider programs managed by others. The reasons for this need exploring, for making tangible progress on harmonisation and reducing the transaction costs of aid to partner governments is crucially important – especially, perhaps, in the context of AusAID's significant role in countries with limited absorptive capacity and in fragile states.

The process of evaluation

90. In terms of the quality and utility of monitoring and evaluation processes, AusAID exhibits commendable emphasis on, and in principle has systems in place for, potentially robust quality and performance monitoring. (These extend well beyond the focus of this study.) However it has to be said that, as currently configured, the utility of its principal, regular, evaluation instrument – the Independent Completion Report – is compromised in a number of ways. Levels of compliance are low resulting in small sample sizes for such a diverse program; the interpretation of evaluative criteria across ICRs and IPRs is variable and quantitative ratings are inconsistent; and – most significantly from an aid effectiveness perspective – the focus is disproportionately pitched on low-level activity management, process and compliance rather than on outcomes and impacts.

91. This study echoes much of the contemporary discussion around performance management in AusAID in the context of scaling up and the putative new Operational Policy Management Framework: *viz.* that program management units need to 'raise their gaze' on monitoring and evaluation and move up from the M&E of individual activities for 'compliance', process-management and input/output-reporting purposes, to M&E for the purpose of informing the strategic positioning of the aid program to achieve higher-level objectives.

92. (Note that this may or may not yield the immediate-term performance stories which public bodies need to be able to tell: this may be more of a communications challenge.)

93. But this requires different expectations and priorities within the already busy work-loads of Posts and other program management units. It implies a more 'intelligent' role – for both Post and implementing partners – in sector analysis, well connected to an understanding of the role of aid in supporting change in the context of the political economy of development. It implies providing the institutional incentives for implementing partners also to focus on managing for outcomes, not least in terms of what is expected from activities' monitoring and evaluation systems. It implies shifting the focus from ensuring evaluations are done to ensuring evaluations are used.

94. In this regard, what AusAID management units ask IPRs and ICRs to assess and deliver on needs to be clearer and unambiguously focused on aid effectiveness and the contribution activities are making to higher-level, often country-strategy, objectives. Evaluations need to be explicitly set in the context of the Country Situation Analysis (the 'why' of the aid program) and the Statement of Commitment (the 'what' of the aid program), and to test the validity of Delivery Strategies (the 'how' of the aid program).

The extent to which such a more strategic evaluation can be properly conducted while some of those contextual documents are not in the public domain needs examining.

95. It implies institutionalising feedback loops and processes of sense-making such that M&E genuinely informs, and is configured and managed to inform, the strategic management of the aid program. Space needs to be made for that, and that space has to be found by reducing the amount of low-level activity, process and compliance clutter.

96. Allied to that, the concept that ICRs principally validate the implementing agent's Activity Completion Report needs to be questioned while those activity reports are so focused on compliance and process rather than the analysis of outcomes and impacts. The complex incentives established through contracts to report success rather than analyse impact also need examining. Those are probably different functions with different processes.

Relationship to Activity Reports

“The format used [in] reports is not revealing in terms of [AusAID] monitoring or evaluation criteria. There is plenty of informational narrative but the presentation of the report is not analytical – it simply states what has been done and not what it means.”

97. Guidance of what evaluators need to consider needs to be expanded to match these higher-level expectations. For example:

- ‘Relevance’ needs to go beyond whether an activity is ‘in line with’ policies and priorities to examine whether, given the wider context, the activity represented the right thing at the right time to contribute maximally to the achievement of higher-level development objectives;
- ‘Effectiveness’ needs to consider not just whether the activity was managed to ‘do what it was supposed to do’, but how correct the theory of change proved to be and an analysis of what contribution the activity made to higher-level outcomes;
- ‘Efficiency’ needs to go beyond commentaries on management quality to consider value for money and whether the activity was designed and configured (not least through the aid modality and approach selected) to achieve the most impact for the effort and resources allocated: efficiency is a ratio of output over input.
- ‘Gender’ needs to discuss not just the numbers of women included in project activities but the impact on society, from the perspective of gender equality, to which the intervention meaningfully contributed.

98. Evaluations do not examine risk sufficiently broadly or strategically. The take on risk is heavily skewed towards financial risks and the choice of modality and management arrangements in ensuring financial probity. This is, clearly, important. However, the analysis of risk also needs to capture *development* risk: that is the risk that the scope, scale and sustainability of development will be affected by (among other things) that very choice of modality and management arrangements. Financial risk needs to be balanced against development risk, for there are often inverse relationships between the two.

Other things

99. Publishing evaluations needs to be more institutionalised, not just so that they are better available to studies such as this one but because that itself would incentivise more robust identification of evaluation requirements and better quality of reporting.

100. A standard lexicon and guide to [a more subtle] AusAID interpretation of key design and evaluation terminology needs to be promulgated. More precise (and internationalised) use of terms such as 'activity', 'output', 'outcome', 'purpose', 'goal' and 'objective' need be championed, not just so that like can be compared with like at evaluation, but so those who manage or hold a stake in an activity are clear about key concepts such as the manageable interest in an activity, and expectations of immediate, intermediate and longer-term impacts.

101. Guidance on style and length of reports needs to promote more succinct, fit-for-purpose (which itself needs definition) documents that lend themselves better to like-for-like meta-analysis. Without exception require short standalone summaries, key data sheets, lists of acronyms and contents lists. Define required length by word-count not by page-count.

Acknowledgements

I am grateful to Adam Blundell and his colleagues in the Secretariat to the Independent Review of Aid Effectiveness for their considerable efforts in sourcing and providing data and other information for the study, always at short notice. My thanks also go to Terry Green and colleagues at theIDLgroup for immediately making available some of the group's researchers to undertake much of the initial data screening and assembly.

Peter Bazeley

February 2011

Annex 1: Terms of Reference

Study of AusAID Independent Completion Reports

The Australian Government has commissioned an Independent Review of Aid Effectiveness (Terms of Reference attached). The Review will be undertaken by a Panel of five persons, chaired by Mr Sandy Hollway AO.

As part of its review, the Panel will commission studies to assist in the overall analysis of the effectiveness and efficiency of the Australian aid program.

One of the commissioned studies will examine the findings of the approximately 190 Independent Completion Reports (ICR's) of AusAID activities completed over the last three years. The study will likely be included as an Annex to the Review.

Specifically, the study will:

- a. Examine and provide a synthesis of the findings of the ICR's, particularly in relation to effectiveness and efficiency;
- b. Analyse any common trends which can be seen from the ICR's and include any lessons which can be learnt from these activities;
- c. Highlight and summarize any particularly interesting or important ICRs, with striking conclusions or lessons, both positive and negative (say, 10 each);
- d. Provide an overall assessment of the quality of the ICRs; and
- e. Provide an analysis on the extent to which the ICRs capture aid programs across the board, or are skewed towards particular types of activities.

The study will include the liberal use of interesting examples from the ICRs when addressing all of the above.

Methodology and Output

It is expected that the study will take 25 days, consisting of 20 days of synthesis and analysis of all the relevant ICR's, and 5 days to draft a report.

The following key output will be produced:

- (1) A report, addressing the terms of reference for the study, and outlining the findings of the synthesis and analysis.

Timing

The study must be completed and the report submitted by 4 February 2011.

Annex 2: AusAID's Current Template for Independent Completion Reports

Aid Activity Name

AidWorks Initiative Number

INDEPENDENT COMPLETION REPORT

Author's Name and Organisation

Date

Activity Summary

| | | | |
|----------------------------|--|-----------------|--|
| Aid Activity Name | | | |
| AidWorks initiative number | | | |
| Commencement date | | Completion date | |
| Total Australian \$ | | | |
| Total other \$ | | | |
| Delivery organisation(s) | | | |
| Implementing Partner(s) | | | |
| Country/Region | | | |
| Primary Sector | | | |

Acknowledgments

Author's Details

Contents

Executive Summary

>

Introduction*Activity Background**Evaluation Objectives and Questions**Evaluation Scope and Methods**Evaluation Team***Evaluation Findings***Relevance**Effectiveness**Efficiency****Impact***

< To determine whether the activity has produced positive or negative changes (directly or indirectly, intended or unintended). The degree to which the various aspects of impact can be assessed will vary according to the nature and duration of the activity. Whether impact can be assessed, or the way impact can be assessed will need to be determined by the Independent Evaluation Team. Impact will not be rated. >

Sustainability***Gender Equality******Monitoring and Evaluation******Analysis and Learning***

Evaluation Criteria Ratings

| Evaluation Criteria | Rating (1-6) |
|-------------------------|--------------|
| Relevance | |
| Effectiveness | |
| Efficiency | |
| Sustainability | |
| Gender Equality | |
| Monitoring & Evaluation | |
| Analysis & Learning | |

Rating scale:

| Satisfactory | | Less than satisfactory | |
|--------------|-------------------|------------------------|----------------------------|
| 6 | Very high quality | 3 | Less than adequate quality |
| 5 | Good quality | 2 | Poor quality |
| 4 | Adequate quality | 1 | Very poor quality |

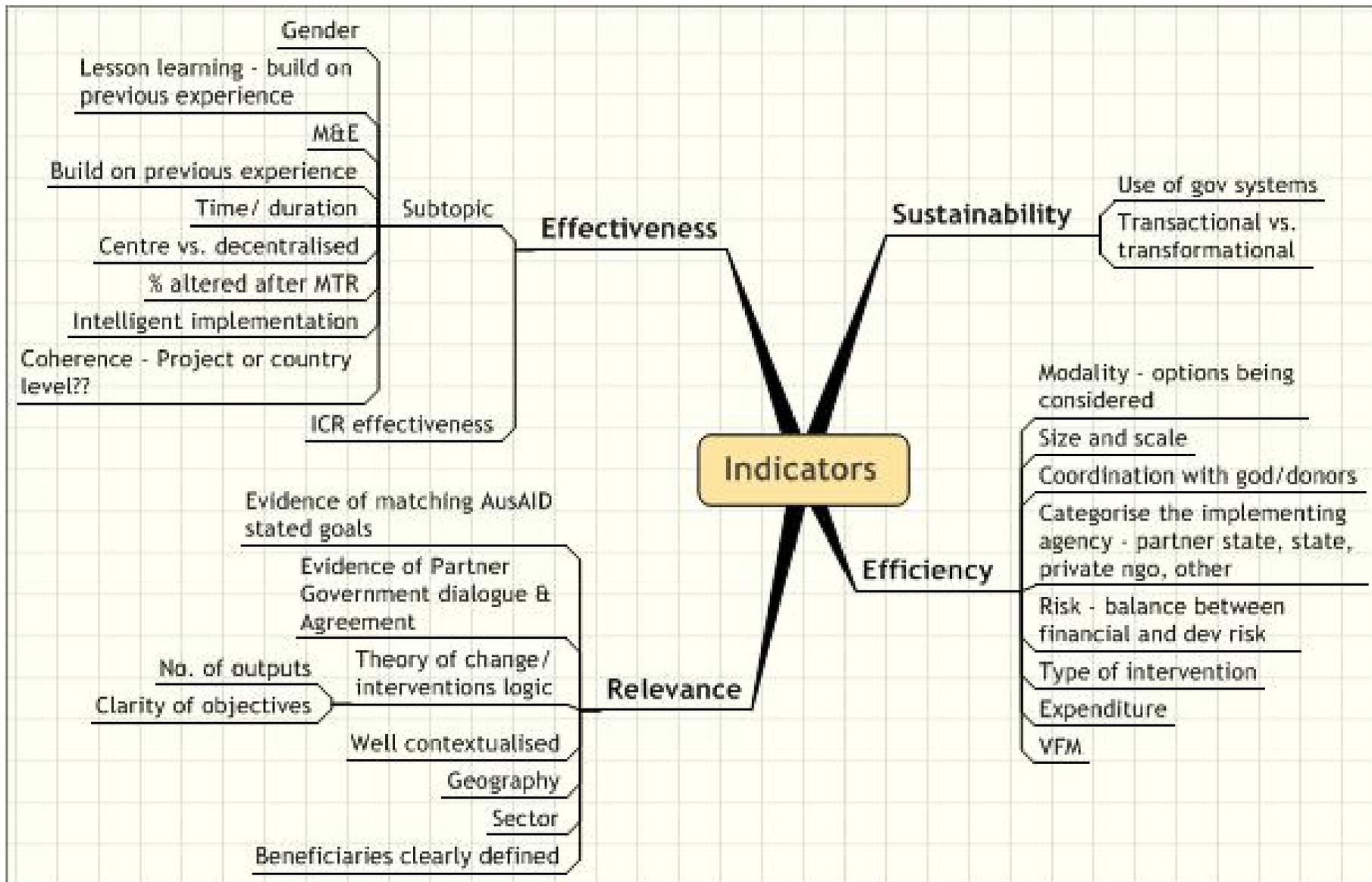
Conclusion and Recommendations

Annex 3: The study's hypothesised 'contributors' to principal evaluation criteria

The study awarded scores to hypothesised 'contributors' to the principal (as regards this review) evaluation criteria of relevance, effectiveness, efficiency and sustainability as follows:

| Scores awarded by this study against selected 'contributors' | |
|--|---|
| Contribution score 1 | Evidence of the contributor playing a significant, positive, part in the 'performance story', and/or it being particularly well managed |
| Contribution score 2 | Evidence of the contributor playing some positive role |
| Contribution score 3 | Evidence of negative contribution and/or being poorly managed |
| Contribution score 4 | Evidence of inappropriate omission from the activity |

The contributors and their hypothesised relationship to aid effectiveness are shown overleaf.



Annex 4: Summarised illustrative activities

The following summaries provide an illustrative selection of some of the ‘best of the best’ and the ‘best of the not so good’ evaluations presented to the study.

- For the ‘best of the best’: activities rated highly at evaluation in (typically) an evaluation which was peer reviewed as also being of good quality;
- For the ‘best of the not so good’: activities rated poorly at evaluation, but for which the evaluation process / report was of good quality.

The best of the best:

The Australia Community Development and Civil Society Strengthening Scheme (ACCESS) Phase 1 in Indonesia

This AUD 22m project was awarded the highest ratings of any activity in the study: all 5’s and 6’s, and 6 (‘very high quality’) overall. Its purpose was to develop and implement an effective approach to sustainable and equitable community-led development within a particular sectoral and geographic focus. It was implemented via a managing contractor.

Key contributors to its quality, as evaluated, included:

- Targeting and supporting ‘enabling environments’ – chasing change;
 - Leading to high levels of ownership by government and community stakeholders;
- Effective engagement with government at all levels;
- The quality of the management team (with a noted ability to recruit and retain high quality staff with the right development skills and values);
- Effective and efficient support by AusAID Post, noting skills and experience, forward planning, and engagement on the ground;
- Leverage of significant civil society, community and government inputs;
- ‘Expert’ championing of established best practice models in community-led development;
- Effective capacity-building strategies;
- An M&E and continuous learning system that effectively captured and fed-back into program design and management results and issues arising.

The program, and therefore the evaluation, was not required to monitor or assess impact on poverty reduction: it was about intermediate outcomes such as civil society more proactively identifying the needs of the poor and marginalised.

The question was raised about the high unit costs of the model adopted and its replicability and sustainability.

Tonga Health Sector Planning and Management Project

This AUD 5.8m managing-contractor-led project improved the Ministry of Health's capacity to plan, manage and deliver health services in Tonga over an eight-year period of implementation. It was one of the highest-rated activities among the evaluations studied, and 'very high quality' overall. While the evaluation conceded that it did not have the information to assess the activity's impact on health outcomes, it considered it "most likely" that health outcomes had been achieved as result of the activity.

Key contributors to its quality, as evaluated, included:

- Strong leadership and ownership by the partner government ministry, with, eventually, most control over project activities placed with the ministry itself;
- A strong, reflective, relationship with the ministry and a demonstrable commitment to "shape and reshape" activities to reflect the ministry's priorities and requirements;
- A solid theory of change and expert-led promotion of international best practice in organisational development and change management;
- A long timeframe in which the focus of activity constantly evolved as capacity developed;

The evaluation also noted that flexibility and responsiveness (which were key to its ultimate success) was initially something Post found difficult to embrace and this caused "much angst to all concerned". However, Post later adopted a much more flexible, adaptive, position and ultimately placed the partner ministry in the driving seat.

The evaluators also noted that the objectives were initially ambiguous and overly ambitious.

Local Governance and Infrastructure for Communities in Aceh (LOGICA) Program

This AUD 28m, three-year, managing-contractor-led project had the goal of making a significant contribution to community recovery and well-being in tsunami-affected areas of Aceh, Indonesia, through the rehabilitation and reconstruction of communities and local government services. Its strategy was to work in three districts to help communities 'decide for themselves and manage what needs to be done', with a focus on re-establishing land rights, sub-district government, village-level communities and community housing.

According to the evaluation, key contributors to its quality – which was principally measured in terms of how well the project delivered its outputs – included:

- Clarity of strategy;
- Capacity-building focused on institutions and systems, as opposed to individuals.
- Focus on transformational rather than transactional activities (even in an essentially humanitarian, post-disaster, context);
- Collaborative design;
- Strong relationship-building;
- Intelligent, flexible, implementation, responding to lessons learned as implementation progressed.

Notwithstanding the evaluators' positive comments on the design itself, they emphasised that the success of the project had more to do with the quality of the collaborative design process, an outstanding team of project staff and excellent working relationships between the project, AusAID and the stakeholders in Aceh.

The evaluation also emphasised how success was more limited in the absence of strong local leadership.

The evaluation critiques the fogginess of objectives, the initially far too short a timeframe (18 months), and limited attention paid to M&E. ("The ICR team came across no evidence that the project had done anything which was not helpful" was the best they could say with any certainty about impact.)

Capital Aid Fund for Employment of the Poor (CEP) Microfinance Program, Vietnam

This AUD 6.3m grant to a Vietnamese microfinance NGO, spread over seven years, was principally used for on-lending but also contained a capacity-building component. The evaluation rated the program's effectiveness, efficiency and sustainability very well. The activities were managed and implemented by the NGO itself: there was no managing contractor.

Key contributors to its quality, as evaluated, included:

- Clear, achievable, objectives with a well-articulated vision and mission;
- Selection of an expert partner organisation that was already and demonstrably highly committed to the objective of poverty reduction, demonstrating excellent leadership, and which was already striving to adopt and develop international best practice in the delivery of micro-finance;
- Effective targeting towards the poor and the poorest;
- A strong transformative 'multiplier' effect in that the capacity developed with AusAID support encouraged other lines of funding to be made available, such that by the end of the program the savings mobilised exceeded AusAID funding by almost 45%.

The evaluators highlighted 'a great deal of anecdotal evidence and comment' to support the view that the program was having an important impact on poverty in its geographical area (Ho Chi Minh City) beyond the immediate impact of the individual loans provided, but conceded that it was not possible to evaluate this empirically from the M&E data available.

Notwithstanding rating sustainability at 5.25 (somewhere between 'good' and 'very high' quality), the evaluators raised some concerns over sustainability, principally because of rising interest rates.

Laos-Australia Basic Education Project

This AUD 8m project sought to improve the relevance, quality and efficiency of primary education, especially for girls, in selected remote ethnic minority areas of Laos – through curriculum and materials development, teacher education and mentoring. It was effectively a discrete grant-funded component of a larger Asian Development Bank (ADB) loan project and was managed and implemented over eight years by a managing contractor. The project exceeded its targets, albeit measured at the level of outputs delivered. It was rated 6 ('very high quality') for relevance, effectiveness and efficiency, and 5 ('good quality') overall.

The impact was deemed principally to be one of demonstrating an effective way of providing teaching in difficult social and geographic contexts, which it did well.

Key contributors to its quality, as evaluated, included:

- Adding an important transformational element to a larger (loan funded) transactional project;
- Close correlation with the partner government's explicit policies and priorities;
- High level of support from the ministry owning the project;
- The flexibility shown between AusAID, the ADB and the managing contractor in working together to manage a complex project with minimal international and national technical assistance.

However, the evaluation noted how opportunities were lost through AusAID too readily adopting the design, and levels of resourcing for the project, previously agreed by the ADB. This impacted on the availability of funding and dropping one important component of the project altogether (relating to quality assurance).

The evaluation also raised concerns about sustainability, in that the improved systemised and approaches still, at the end of the project, needed to be institutionalised ... and funded at scale.

Nusa Tenggara Timur Primary Education Partnership

This AUD 27m, seven-year, capacity-building program aimed to improve education service delivery in three districts in one province of eastern Indonesia. The activity, implemented through a managing contractor, involved all 936 schools across the three districts, but with a focus on 214 'Partner Schools'. It had identified low quality teaching and learning, low retention to secondary education, community disenchantment with education and a perceived irrelevance of education.

The activity was rated very highly at evaluation, which found that it had achieved significant progress in working with stakeholders to achieve the objectives of the partnership: measurable and significant improvements to teaching and learning (not least for girls), more effective schools-based management, and increased community participation.

Key contributors to its quality, as evaluated, included:

- The development of "a deep understanding at systems level" of the drivers of improved student outcomes;
- A focus on key transformative aspects: the systems, processes and governance arrangements (including stakeholder participation) that promote continuous improvement in teaching and learning;
- Working closely with stakeholders;
- Building on existing structures and processes;
- Expert-led adoption of proven models of best practice;
- 'Very close' alignment with explicit government policy and sector-development priorities;
- Increased political will as the partnership developed with – now – enhanced government budgetary allocations, through concerted engagement with local government.

The evaluation also benefitted from good baseline data on which to assess performance.

Mozambique In-Country Scholarship Assistance Scheme (ICSAS)

The goal of this AUD 1.6m (over five years) scheme was to assist academically-qualified people from the provinces and from poor and disadvantaged families to gain university and other tertiary qualifications in specific fields. Its immediate objective, or purpose, was to establish a funding mechanism whereby 300 Mozambican nationals would receive Australian financial assistance.

The activity “attained its objectives overall” and was deemed to be of ‘good quality’ in all aspects, and to have been ‘very high quality’ in terms of effectiveness and value for money.

It was managed in-house by AusAID.

Key contributors to its quality, as evaluated, included:

- Appropriate choice of delivery partner;
- “Some correlation” with the government’s own scholarship program;
- Being “genuinely demand-driven”;
- Effective response to “serious management and operational shortcomings” identified at an earlier review.

However, the evaluation highlights how even the higher- (goal-) level objective was simply to undertake a process – to provide scholarships – and that it was therefore neither the project’s nor the evaluation’s responsibility to assess the impact that those scholarships might be having on national development.

The evaluation also highlighted the need for greater alignment and donor harmonisation in the scholarships sector, and for better targeting around identified national needs and workforce requirements.

❖ *In the view of this study, this is an example of how poorly-defined objectives and the definition of effectiveness being pitched at the level of activity and process can lead to high ratings at evaluation where they cannot – from an aid-effectiveness perspective – really be justified. Indeed, the evaluation itself catalogues shortcomings that might suggest that the scheme did not, in fact, necessarily result in the development outcomes that were presumably implicit in design, although it notes that there was no provision for post-scholarship monitoring.*

The best of the not so good:

Fiji Education Sector Program

This AUD 25m program, really a standalone project, aimed to assist the Ministry of Education to implement strategic reforms, thereby improving the delivery and quality of education in the country. Every aspect of the program was rated less than satisfactory, with relevance and sustainability rated as 'poor'.

The evaluation noted "numerous reasons to be concerned".

Key contributors to its disappointing contribution to aid effectiveness, as evaluated, included:

- Standalone implementation arrangements, distanced from government systems;
- Disconnection with government planning and budgeting;
- Technically unsound / lacking best practice (for example in the area of curriculum development);
- Erosion of the program's relevance as the context changed but the program didn't;
- Suboptimal approaches to the provision of technical assistance, annual planning, supervision, reporting and activity management;
- Blindness to critical transformational issues (such as, for example, teachers' pay and conditions);
- Little empirical evidence of results and adapting in response to progress.

Media for Development Initiative: Australian Broadcasting Corporation / National Broadcasting Corporation Project Phase 2, Papua New Guinea

This twinning arrangement (of unstated cost) between the Australian Broadcasting Corporation (ABC) and PNG's National Broadcasting Corporation (NBC) was "to contribute to strengthening of NBC's organisational capacity as the national broadcaster" so as "to strengthen the NBC to enable it to improve the quality, distinctiveness, universality and efficiency of its programs".

Outputs were to "improve", "assist" and "encourage" internal processes at NBC.

The evaluation concluded that the "overall effect or outcome" of the project was "a renewed optimism within the NBC that it can fulfil its role". This is based largely on the completion and endorsement of a new corporate plan. The review found "a strong, professional and successful partnership between AusAID, the ABC and the NBC" and recommended its continuation. It recognised strong leadership and ownership shown by NBC staff.

However, actual delivery of broadcasting services was a major concern, as was progress on a number of technical improvements, and the project attracted the lowest evaluation ratings of any in the study.

Key contributors to its [poor] quality, as evaluated, included:

- Fragmented project management, resulting in inadequate planning and reporting, with confused roles and responsibilities;
 - Lack of oversight and governance arrangements;
 - Lack of functional M&E systems, and lack of learning;
 - "Very little understanding within the project team of the need to actively program for sustainable outcomes";
 - Ditto regarding knowledge of development good practice such as "non-directive consulting techniques" and "attention to operating within government procedures and regulations to ensure sustainability".
 - Advisers providing "seemingly unstructured advice and informal on-the-job training on a day-to-day basis";
 - Disproportionate focus on quick wins to the detriment of skills and sustainability;
 - Change without the evidence of its need or identification of the problem;
 - A misplaced assumption by AusAID that ABC had the understanding of, and capacity to meet, AusAID's design and reporting requirements.
- ❖ *In this study's view, this was a classic case of a supply-led project with foggy objectives based on no clear theory of change, implemented without sufficient attention to the fundamental elements of good development practice.*

**Papua Public Expenditure Analysis & Capacity Enhancement (PEACH),
Indonesia**

Earlier World Bank analysis had suggested that weak public financial management (PFM) was the main factor behind ineffective resource allocation in Papua, despite large fiscal flows. The World Bank-managed PEACH program firstly undertook public expenditure analysis and then, secondly, strengthening of PFM capacity. AusAID financed that second component to the tune of AUD 1.3m over 2½ years.

The evaluation considered that the project would not achieve its stated outputs (sic) and that its overall impact would be limited. The project was rated as less than satisfactory in all categories except 'effectiveness' (curiously), and was rated 'poor' for sustainability and gender.

Key contributors to its poor quality, as evaluated, included:

- Significant early design flaws inherited, but not influenced (or peer reviewed) by AusAID, resulting in an overly narrow, technical, focus;
- Overly ambitious objectives: too many to be realistically monitored and evaluated;
- Located in a policy vacuum;
- It required linkages to national policy, which were not there;
- Limited leverage with provincial government;
- Technical team championing reforms that were ahead of the partner government's position and willingness and capacity to adopt;
- Under-resourced, with too short a timeframe, and unable to respond to identified needs essential to sustainability.

The evaluation did however note flexibility in implementation, which went some way to mitigating the impacts of poor design.

Framework for Australian Assistance to Disaster Management in Papua New Guinea

This 2008 evaluation examined the quality of an AUD 8.8m AusAID contribution over five years, the [revised] goal of which was 'an enhanced and more integrated national network for disaster risk management in PNG', and its purpose was 'more effective and timely support for disaster mitigation, preparedness, response and recovery at national, provincial and community level'.

The 'framework' was essentially delivered through a number of separately managed and administered activities, overseen by AusAID Canberra (rather than Port Moresby) until mid-2006. The project includes (among other things) twinning arrangements with an Australian whole-of-government partner and an Australian NGO.

The evaluation was conducted as a part of the design of a new, follow-on, program.

Although the cooperation was subsequently extended, the evaluation rated the activity as less than satisfactory overall, considering that the framework had had limited success in achieving its objectives, with only marginal increase in the capacity of agencies and organisations in the sector and 'questionable' sustainability.

Key contributors to its low quality, as evaluated, included:

- Unrealistic objectives;
- Both the partner government and AusAID had other 'more pressing' priorities, with disaster management not a high priority in its medium-term development strategy;
- The design did not reflect local capacity;
- AusAID's 'modest' level of involvement in the sector;
- The 'modest' response to the needs of the sector, focused on only selected sub-sectors, and the project's short timeframe;
- The [appropriate] diversion of attention to a major cyclone-related disaster, which occurred mid-project. (The response to which, it was noted, almost certainly benefited from the project's recent training.)

The evaluation noted the relevance of funding the United Nations Office for Coordination of Humanitarian Affairs.