



# Review of Operational Evaluations completed in 2014

## Part 3: Annexes

### Annex A: Review methods

#### 1. Introduction

The Review of Operational Evaluations examined a sample of 35 independent operational evaluations finalised in 2014. It aimed to assess the quality of these evaluations and facilitate opportunities to learn from their findings. This Review follows the first Review of Operational Evaluations which examined operational evaluations completed in 2012.

This Annex outlines the methods for the Review. It was undertaken by staff from the Office of Development Effectiveness (ODE). It was a relatively modest undertaking which aimed to provide findings which are useful but only required a moderate investment of ODE staff time.

As a general principle, there were minimal differences between the methods used in the first review and the current Review. This Review was predominately be a desk review of completed operational evaluation products. It included the identification of evaluations which provide useful learning on aid activities for DFAT and the international development community. The desk review was complemented by a short all-staff survey on DFAT's evaluation capability and an internal audit on the publication of operational evaluations.

#### 2. Objectives

The Review has three objectives:

- a. To better understand the practices related to, and the quality of, independent operational evaluations and how these have changed since 2012
- b. To provide information to support good quality, independent evaluations across the department; and
- c. To promote better use of evaluations across the department and the aid community by facilitating opportunities for learning.

#### 3. Scope

When scoping for the Review commenced, 66 independent operational evaluations that were completed in 2014 were identified.\* This is more than the 55 evaluations which were identified when planning for the Review began. Given the increase in numbers, it was proposed that a sample of 36 evaluations be examined to ensure the Review can be managed within the available capacity of the review team.

---

\* As at 30 April 2015. At the end of July 2015 a total of 77 evaluations completed in 2014 had been identified. The number of evaluations has increased because there is often a lag time between completion of an evaluation and when it is uploaded into Aidworks.

During the assessment process, it was found that one document included in the sample was a piece of research, rather than an operational evaluation. This document was therefore excluded from the Review, reducing the sample to 35 operational evaluations.

A purposeful sampling strategy was used. This ensured that sectors, divisions and country programs were adequately represented in the evaluations examined. Approximately half of evaluations from each division, and approximately half of evaluations from each sector, were included in the sample. In addition, at least one evaluation from each country program was included. Please see Appendix B for an outline of the sample used\*.

Because a purposeful sample was used, from a statistical perspective care is required if seeking to generalise the findings to all 2014 evaluations. However, the Review findings still provide useful information on the characteristics and quality of the evaluations which are examined.

## 4. Audience

There are two primary audiences for this Review: staff from ODE and performance and quality (P&Q) staff from DFAT's program areas. The most relevant evaluation questions for both of these groups, and how the evaluation findings can be used by them, are outlined in Appendix A.

The key secondary audiences for the Review are the DFAT Executive and senior managers from the Contracting and Aid Management Division (ACD). The findings from the Review can be used by these groups to inform decisions on the department's investment quality reporting system.

Aside from these groups, all DFAT staff involved in commissioning and managing evaluations will have an interest in the Review's findings. This is because the findings will assist staff to commission and manage higher quality evaluations. All DFAT aid staff and the broader aid community will also have an interest in the opportunities for learning which are provided by the Review.

The findings of the Review will be shared with DFAT staff and other stakeholders in the following ways:

- › The Review report will be published on the ODE website
- › A four-page summary will be printed and distributed at appropriate forums (see below)
- › ODE staff will present the Review process and findings at appropriate DFAT forums, such as Performance and Quality Network meetings, and
- › If opportunities arise, Review findings may also be presented at external forums such as the Australasian Evaluation Society's annual conference.

---

\* Random sampling was also considered. It was not chosen as, given the small number of total evaluations, a random sampling approach may (by chance) result in a sample of evaluations which does not include important sectors or country programs.

## 5. Evaluation questions and methods

The Review addressed the following evaluation questions:

---

### Priority questions

1. What are the characteristics and quality of operational evaluations? How have these changed since 2012?
2. What factors contribute to the quality of operational evaluations?
3. To what degree do operational evaluations provide a credible source of evidence for the effectiveness of the Australian aid program?

### Other key questions

4. What is the capability to commission and conduct operational evaluations across the department?
  5. How are the findings from operational evaluations used in the department?
  6. How can the publication rate of operational evaluations be improved?
  7. How do DFAT's evaluation policies and practices compare to the evaluation policies and practices of other donors?
- 

Questions 1-3 were identified as high priority. This is because they focus on the quality of operational evaluations. Evaluation quality is of primary importance to this Review because it is a pre-requisite for evaluation use, learning and publication. If an evaluation is not considered to be credible, it is unlikely its findings will be used and it may not be published.

The other key questions (4-6), while important, would require a high level of resourcing to address in a comprehensive fashion. As such resources are not available, the methods used to address these questions was limited.

The following section outlines how the quality of operational evaluations was assessed in the Review. The data collection and analysis methods for each question are then described (see Appendix A for a summary table).

### Assessing evaluation quality

The most resource intensive task in this Review was to assess the quality of each operational evaluation. This assessment provided the data needed to address many of the Review questions.

The quality of each operational evaluation was assessed using the pro-forma at Appendix C. This pro-forma includes nine criteria which are based on DFAT's Monitoring and Evaluation Standards. The criteria in the pro-forma therefore represent what the department believes should be included in good quality evaluation products (comprising evaluation reports, terms of reference and evaluation plans). A similar approach to assessing evaluation quality was successfully used in the previous Review and has been used in meta-evaluations conducted by other donors.

The pro-forma is largely the same as the pro-forma used for the first Review of Operational Evaluations. This is to ensure reasonable comparisons can be made between the evaluations assessed in each Review. However, there are three differences between the current pro-forma and the previous pro-forma:

1) The first Review examined operational evaluations conducted in 2012. At that time evaluations were required to assess investments against six aid quality criteria: relevance, effectiveness, efficiency, sustainability, impact and gender equality. Therefore the first Review pro-forma included an assessment of how well evaluations assessed investments against these criteria.

However, it is no longer compulsory for evaluations to assess investments against these criteria and, as a result, they have been removed from the pro-forma. Note that the Review will still collect data on whether an evaluation uses the criteria (see pro-forma) but will not consider the quality of assessments against the criteria.

2) The first Review pro-forma included the criterion 'assessment of intervention logic'.\* This criterion has been removed from the pro-forma for this Review because the department's most recent evaluation guidance, which was introduced in mid-2014, emphasises that evaluations should be 'fit-for-purpose'. This means they should focus on the information needs of investment managers. A high quality assessment of intervention logic is an in-depth task that may not meet the needs of an investment manager. Further, the department's M&E Standards and evaluation guidance do not require, or even suggest, that operational evaluations must assess an investment's intervention logic. As a result, this criterion has been removed.

3) An additional criterion of 'Executive Summary' has been added. This criterion will be used to assess whether an operational evaluation report's executive summary provides all the necessary information to enable primary users to make decisions. Experience tells us that the executive summary is a very important part of an evaluation report, as it is often the only part senior staff have time to read. The inclusion of this criterion is consistent with DFAT's Monitoring and Evaluation Standards and good practice for long reports.

## Priority Questions

### **Q1: What are the characteristics and quality of operational evaluations? How have these changed since 2012?**

This question aimed to build an understanding of the types of investments being evaluated and the nature of these evaluations. A range of information was collected, for example:

- › Investment value, sector and country/region
- › Evaluation cost and duration
- › Whether evaluations were partner-led, joint, or DFAT-led.

Analysis was then conducted to produce a range of descriptive statistics for investments and their evaluations, for example:

- › The number and percentage of evaluations conducted by sector and country/region
- › Average cost and duration of evaluations
- › The number and percentage of evaluations which are partner-led, joint and DFAT-led.

A full list of characteristics collected, the source of data and analysis conducted is at Appendix D.

As outlined above, the quality of operational evaluations was assessed using the Review pro-forma. The assessment results were recorded in a central database and each criterion was analysed to identify specific areas where evaluation quality was high or low.

---

\* This criterion was fully described as "The evaluation assesses the intervention logic or theory, or equivalent, including underlying assumptions and factors. The report assesses the clarity of initiative objectives".

A measure for overall evaluation quality was also established. In the first Review the pro-forma criterion 'credibility of evidence and analysis' was used as a proxy for overall evaluation quality. This is because this criterion was most strongly associated with the other quality criterion in the Review.

For this Review, correlation analysis of the nine pro-forma criteria was conducted to establish whether 'credibility of evidence and analysis' was still the best predictor of evaluation quality. This criterion had high correlation with the other criteria. Other criterion demonstrated similarly high correlation, however, it was decided to retain 'credibility of evidence and analysis' as the proxy for overall evaluation quality. This was because:

- › This criteria focuses on the logical flow and evidence base on an evaluation. Using a common sense test, this is a good marker of overall evaluation quality
- › Using the same measure of overall evaluation quality as the previous Review allows for easier comparison between Reviews.

Based on the overall measure of evaluation quality, broad conclusions were drawn on the quality of operational evaluations. The overall measure of evaluation quality was also used to address evaluation question 2 (see below).

Investment and evaluation characteristics, and evaluation quality, were compared to the results of the previous Review through a series of tables and figures. Analysis focused on areas where the characteristics or quality have changed substantially since the previous Review.

## **Q2: What factors contribute to the quality of operational evaluations?**

The aim of this question was to understand the factors that make a strong contribution to evaluation quality. Identifying such factors will allow DFAT staff to focus on a small number of critical areas that contribute to high quality evaluations.

The first Review, as well as similar meta-evaluations commissioned by other donors, have considered a number of factors which may contribute to or drive evaluation quality. However, findings of different studies have been mixed. For example, the first DFAT Review found a positive relationship between evaluation duration and evaluation quality. An equivalent USAID meta-evaluation (2013) surmised that a relationship exists between evaluation quality and evaluation duration/cost, but did not have sufficient data to support this. A SIDA meta-evaluation (2008), on the other hand, found no clear correlation between evaluation cost and quality (see Appendix E for further information).

This Review identified six factors from the first Review and similar meta-evaluations where there is some or mixed evidence about the relationship between the factor and evaluation quality. These factors are:

1. Duration of an evaluation (total days and field days)\*
2. The purpose of an evaluation (e.g. progress or completion evaluation)
3. Evaluation team size
4. Evaluation team composition, particularly the presence of specialist evaluator skills
5. Quality of investment M&E systems
6. The number of evaluation questions addressed.

As noted under Q1 above, an overall measure of evaluation quality was established. Data on factors 1-4 was collected using the Review pro-forma. Additional data was collected on factor 5 (quality of investment M&E systems) and factor 6 (the number of evaluation questions). This data was easily obtainable. The quality of

---

\* Note evaluation cost is seen as a proxy for evaluation duration, and so has not been included as a standalone factor

investment M&E systems was obtained from the department's aid quality checks (AQC), which are annual self-assessments of investment quality. The number of evaluation questions were identified as part of the pro-forma assessment process.

Correlation analysis\* was initially conducted to examine the relationship between the above six factors and evaluation quality. However, correlation analysis did not identify any significant relationships. Further analysis was conducted and, through this, non-linear relationships between evaluation quality and the six factors were identified (the non-linear nature of the relationships explains why correlation analysis did not provide useful results). Note that both correlation and other analysis found no relationship between evaluation quality and the quality of an investment's M&E system.

Given the small number of evaluations examined, the Review needs to treat the findings cautiously when making inferences from the above analysis. However, the results of the analysis, combined with the findings of other meta-evaluations, will assist DFAT to identify a small number of critical areas that can contribute to high quality evaluations.

Note that a number of other factors that contribute to evaluation quality have been identified through other meta-evaluations. These include the quality of the evaluation design; the presence of a strong relationship between the commissioning agency and the evaluation team; and the evaluation capacity of staff in the commissioning agency. However, the relationship between such factors and evaluation quality were not examined in this Review. This is because significant resources would be needed to collect and analyse the qualitative data required to understand these factors, which was beyond the capacity of the Review team.

### **Q3: To what degree do operational evaluations provide a credible source of evidence for the effectiveness of the Australian aid program?**

The aim of this question was to understand the extent to which operational evaluations are a credible source of evidence for the Australian aid program. This was addressed by examining the assessments against pro-forma criterion 8, 'credibility of evidence and analysis'. The findings from these assessments were used to draw conclusions about the credibility of operational evaluations as a source of evidence. Note that, given conclusions were drawn from this single source of evidence, their strength is modest.

### **Other key questions**

The four questions below were medium priority questions for the Review. To address these questions comprehensively would require extensive data collection and analysis, with a particular focus on qualitative data. This was beyond the resources available for the Review. As a result, each of these questions was addressed using limited data collection and analysis techniques. Where possible, triangulation was undertaken by comparing the findings of this Review with other similar meta-evaluations.

### **Q4: What is the capability to commission and conduct evaluations across the department?**

The aim of this question was to better understand the capabilities of DFAT staff to commission and conduct operational evaluations. This can be used to guide ODE advice and Executive decision making on the department's capability development needs.

To address this question, data was collected using a short electronic survey. The survey focused on staff training, skills and experience in commissioning evaluations, participating in evaluation teams, and leading evaluation teams. A survey is at Appendix F.

---

\*The chi-square correlation coefficient was used, given the data was categorical (rather than numerical). Note that multiple regression analysis was also considered, however, the Review's data set is too small to produce meaningful results.

The survey was made available to all DFAT staff. Targeted promotion and follow-up with staff who have an interest in evaluation, such as members of the Performance and Quality network and informal networks was also undertaken. Given the ‘survey fatigue’ within the department and the specialised nature of the survey, it was anticipated the response rate would not be high and that most respondents would have an interest and experience in evaluation. We were pleasantly surprised when 113 staff responded to the survey.

Survey results were analysed to identify the number of staff with skills, training and experience related to commissioning and conducting evaluations.

#### **Q5: How are the findings from operational evaluations used?**

The aim of this question was to develop a better understanding of how evaluation findings are used in the department. To do this, management responses for evaluation reports\* were examined to determine the number and percentage of evaluation recommendations which have been accepted, partially accepted, or not accepted. Based on these findings, modest conclusions were drawn on evaluation use.

#### **Q6: How can the publication rate of operational evaluations be improved?**

DFAT’s policy is that all operational evaluation reports should be published on the DFAT website. However, the publication rate has been low for a number of years. This is not only of concern to ODE; DFAT’s Internal Audit Branch (AUB) had signalled its intent to examine this issue in its forward work plan. ODE and AUB agreed that the two areas would collaborate on an investigation into DFAT’s evaluation publication practice. AUB conducted an internal audit on evaluation publication in late 2015. The audit report, which was made available in early 2016, was used to inform the findings of the Review.

#### **Q7: How do DFAT’s evaluation policies and practices compare to the evaluation policies and practices of other donors?**

DFAT’s evaluation policies and practices were compared with those of similar bilateral aid agencies. The main source of data was the OECD Review of Evaluation Systems in Development Cooperation. This review, which is partly funded by ODE, examined the evaluation systems of DAC members, including their evaluation policies and mandates; planning and decision making with regard to what to evaluate and how; and implications for evaluation of mergers between development and foreign policy ministries. The draft report became available in early 2016.

### **Analysing data across evaluation questions**

To ensure the Review provided information to support good quality independent evaluations across the department, the findings from the evaluation questions were brought together to identify the implications for DFAT’s evaluation policies and practices. Analysis focused on whether DFAT’s current evaluation policy encourages intended practices in the department, potential areas where DFAT’s evaluation policy could be adjusted, and lessons that could be adopted from other donors.

### **Opportunities for learning**

As noted above, one of the Review’s objectives was to promote better use of evaluations by facilitating opportunities for learning. This is because operational evaluations may provide sectoral or evaluation lessons which are useful for DFAT staff and the broader aid community.

The first step in meeting this objective was to identify operational evaluations which meet a minimum quality threshold. This was necessary because an evaluation needs to be credible if its findings are to be broadly used. Using the overall measure of evaluation quality which is discussed under evaluation question 1, a list of sufficiently

---

\* Note it is anticipated that management responses will not be available for all evaluation reports. This is because some management responses will not have been completed or because it will not be easy for the ODE team to locate them.

credible evaluations was developed. Four approaches have been implemented to facilitate learning from these evaluations.

First, when assessing operational evaluations, team members identified good practice evaluation products (terms of reference, evaluation plans, evaluation reports and management responses). These are products which performed highly against the criteria outlined in the Review pro-forma and, as a result, provide DFAT staff and partners with concrete examples of high quality evaluations. Five to six good practice products were identified covering different sectors and countries/regions. These good practice products are included in the Review report and will be made available on the ODE website.

Second, team members identified operational evaluations which provide useful learnings on aid and development for DFAT staff and development partners. Such evaluations were identified using a set of criteria, for example, evaluations which use innovative methods; evaluations which credibly demonstrate the achievement of higher order outcomes; and/or common themes that occur in several operational evaluations based on the same sector (see Appendix C for the full list of criteria). Extracts or summaries from five to six evaluations are included in the Review report as short 'evaluation snapshots'.

Third, lessons from operational evaluations will be used to inform ODE's planned strategic evaluations (for example, economic infrastructure or aid for trade). Any operational evaluations which meet the quality threshold and which are relevant to an ODE strategic evaluation will be provided to the relevant ODE evaluation manager. This manager can then use the operational evaluations to inform the strategic evaluation.

Fourth, the list of operational evaluations which meet the quality threshold will be sorted by sector (for example, governance, health, education etc). The list has been provided to the Development Policy Division and the Aid Contracting and Management Division for their use in assessing the quality of the overall aid program. Note that ODE will not have ongoing involvement in this work.

## 6. Resources

As noted in the introduction, the Review was conducted by ODE staff. Table 2 outlines the team and roles.

All team members were asked to record the approximate time they spent working on the Review. This allowed ODE to better understand the staff time and resources required to implement the Review.

A small budget of approximately \$2000 was required for the Review. This will predominately be used for finalising documents and printing Review outputs such as a four page summary.

**Table 2: Review roles and responsibilities**

Review position	ODE position	Key roles in Review
<b>Team Leader</b>	Director, Operational Evaluations Section	Accountable for the Review. Tasks include: <ul style="list-style-type: none"> <li>&gt; Overseeing Review process</li> <li>&gt; Ensuring consistency of assessments across the team</li> <li>&gt; Assessing the quality of operational evaluations</li> </ul>
<b>Review Manager</b>	Assistant Director, Operational Evaluations Section	Coordinate day to day Review tasks, including: <ul style="list-style-type: none"> <li>&gt; Creating data recording and management systems</li> <li>&gt; Assessing the quality of operational evaluations</li> <li>&gt; Coordinating team members' input</li> </ul>



		<ul style="list-style-type: none"> <li>&gt; Conducting data analysis</li> <li>&gt; Coordinating and drafting key documents with input from other review team members</li> </ul>
<b>Review Assistant</b>	Administrative Officer, Operational Evaluations Section	Responsible for data management, including: <ul style="list-style-type: none"> <li>&gt; Collecting operational evaluation documents</li> <li>&gt; Recording basic characteristics of evaluation</li> <li>&gt; Data entry</li> <li>&gt; Research tasks as needed</li> </ul>
<b>Other team members</b>	Assistant Director, OES  Director, EVA  Assistant Director, EVA	<ul style="list-style-type: none"> <li>&gt; Assessing the quality of operational evaluations</li> <li>&gt; Contributing to data analysis and drafting key documents</li> <li>&gt; Assistant Director OES will manage data collection and analysis for evaluation question 5</li> </ul>

## 7. Limitations

One limitation is that ODE will be assessing the evaluation policy, guidance and support which it oversees and provides. This limitation was addressed as follows:

- > The Independent Evaluation Committee (IEC) oversaw the Review process and findings. This helped ensure any self-assessment conducted by ODE is defensible, and
- > The limitation is clearly acknowledged in the evaluation report to ensure readers take it into account.

Another limitation is that ODE will be required to respond to the Review's findings. It is difficult for ODE to both draft recommendations and a management response to the recommendations. To address this, the Review report includes recommendations that will be considered by the IEC. The IEC will note whether the recommendations flow logically from the report and are supported or not. If the IEC supports the recommendations, ODE will act on them. No formal management response will be completed.

The Review included the evaluation 'Syria Crisis – Humanitarian Assistance'. The Director of ODE's Operational Evaluation Section was/is the team leader for both the Syria evaluation and this Review. The evaluation was included in the Review because it is the only evaluation for the Middle East program. The following steps were taken to counter any real or perceived conflicts of interest:

- > The quality of the Syria evaluation was assessed by either the Review team leader or the members of his section (who are all involved in the Review). Rather, it was assessed by the Assistant Director EVA. This ensured the evaluation's quality was assessed in an objective manner
- > The Syria evaluation was not eligible to be a good practice evaluation product. This was to counter any perceptions that ODE was overly favourable when assessing evaluations produced by ODE staff.

### Consistency of assessments across the team

The quality of operational evaluation reports was assessed by five team members using the pro-forma at Appendix C. To ensure the findings of the Review were credible, it was important to ensure team members assessed operational evaluations relatively consistently. This was achieved as follows:

1. A Review handbook was created which provides a short description of each criterion, including what 'adequate' quality for each criterion looks like.

2. Prior to starting assessment of operational evaluations, the Review team met and discussed the assessment pro-forma and handbook. This provided a common understanding of the criteria and the ratings to be used.
3. As the first step in the assessment process, all team members assessed the same operational evaluation using the pro-forma. The Review team then met and discussed the assessments. This further built a common understanding of how to assess the evaluations and identified any criteria where further clarification or changes to the pro-forma were required.
4. Part way through the assessment process, Review team members again assessed the same operational evaluation and met to discuss their assessments. This was to further check that all team members were conducting consistent assessments.

## 8. Ethical conduct

Consistent with ODE policy, the Review was guided by the AES Guidelines for the Ethical Conduct of Evaluations.

Issues around ODE's assessment of its own policies, guidance and evaluations are discussed under 'Limitations' above.

## Appendix A: Summary of evaluation questions and methods

Evaluation question	Primary intended users	Data collection methods	Data analysis methods
<p><b>1. What are the characteristics and quality of operational evaluations? How have these changed since 2012?</b></p>	<p>ODE: to better understand current practices in the department and to provide relevant advice to the Executive</p>	<p>Basic characteristics of 2014 evaluations (and the investments they relate to) collected from evaluation database, AidWorks and Review pro-forma</p> <p>2014 evaluations rated against quality criteria in Review pro-forma</p> <p>Evaluation characteristics and quality summarised from the first Review of Operational Evaluations</p>	<p>Descriptive statistics to be derived, such as number, average and range for evaluation and investment characteristics</p> <p>Analyse each pro-forma criteria to establish areas where evaluation quality is high and low.</p> <p>Establish a measure to assess overall evaluation quality.</p> <p>Comparative tables/ figures of evaluation characteristics and quality, comparing the first and current Reviews of Operational Evaluations</p> <p>Analysis of comparatives tables to assess changes to evaluation practices.</p>
<p><b>2. What factors contribute to the quality of operational evaluations?</b></p>	<p>ODE and P&amp;Q staff: to provide advice to program staff on factors to focus on when planning evaluations</p>	<p>Review of other meta-evaluations to identify possible factors contributing to evaluation quality</p> <p>Data on such factors and evaluation quality collected under Q1 above.</p> <p>Data on quality of investment M&amp;E systems to be collected from Aid Quality Reports.</p>	<p>Correlation analysis to examine relationship between evaluation quality and possible factors contributing to evaluation quality</p>

<b>3. To what degree do operational evaluations provide a credible source of evidence for the effectiveness of the Australian aid program?</b>	ODE staff: to support ODE assessments of the Investment Quality System and the Performance of Australian Aid report	Data on evaluation quality collected under Q1 above.	Analysis of assessments against pro-forma criterion 8, 'credibility of evidence and analysis'.
<b>4. What is the capability to commission and conduct operational evaluations across the department?</b>	ODE: to better understand current practices in the department and to provide relevant advice to the Executive	Short survey on evaluation capability conducted	Quantitative analysis of survey responses
<b>5. How are the findings from operational evaluations used in the department?</b>	ODE: to better understand current practices in the department and to provide relevant advice to the Executive	Management responses assessed to examine which recommendations accepted	Examine percentage of recommendations which were fully, partially or not accepted
<b>6. How can the publication rate of operational evaluations be improved?</b>	ODE and P&Q staff: to improve policies and systems for publication	Work together with the Internal Audit Branch to develop methods to collect and analyse data on why publication rate is low and how it can be improved	
<b>7. How do DFAT's evaluation policies and practices compare to the evaluation policies and practices of other donors?</b>	ODE staff: to inform operational evaluation policies and guidance	To be collected through the OECD DAC Review of Evaluation Systems in Development Cooperation	Analysis to determine pertinent points which will support good quality independent evaluations across the department

## Appendix B: Sample

The purposeful sample aims to include:

- › Approximately half of all evaluations produced by each Division
- › Approximately half of all evaluations per sector, and
- › At least one evaluation per country program.

**Table 1: Purposeful sample by Division and Sector**

The table below outlines the number of evaluations per division and per sector which will be included in the sample.

	DPD	MAD	MPD	OTN	PAD	SWD	SRD	SED	TED	Sample per sector	Total evaluation per sector
Agriculture		1								1	2
Education					1	1		1		3	6
Forestry						1				1	1
Government and Civil Society	3		1		2	2	3	2		13	26
Health					1		1			2	4
Humanitarian Aid		1				1				2	4
Mineral Resources and Mining									1	1	1
Multisector/Cross-Cutting		1		1		2	3			7	14
Other Social Infrastructure and Services					1					1	1
Population Policies/Programmes and Reproductive Health	1						1			2	3
Trade Policy and Regulations and Trade-Related Adjustment							1			1	2
Transport and Storage					1			1		2	2
Sample per division	4	3	1	1	6	7	9	4	1	36	
Total evaluations per division	8	5	3	1	10	13	17	8	1		66

Division key:

**DPD:** Development Policy Division

**MAD:** Middle East and Africa Division

**MPD:** Multilateral Policy Division

**OTN:** Office of Trade Negotiations

**PAD:** Pacific Division

**SWD:** South and West Asia Division

**SRD:** South-East Asia Mainland and Regional Division

**SED:** South-East Asia Maritime Division

**TED:** Trade, Investment and Economic Diplomacy Division

**Table 2: Purposeful sample by country program**

The table below outlines the evaluations per country program which will be included in the sample (note that only Divisions that include country programs are shown in the table).

Division and country program	Total Evaluations	Sample by country program
<b>Middle East &amp; Africa Division (MAD)</b>	<b>5</b>	<b>3</b>
> AFE - Pan Africa	1	1
> AFW - Southern Africa	3	1
> ME1 - Middle East	1	1
<b>Pacific Division (PAD)</b>	<b>10</b>	<b>6</b>
> FJ1 - Fiji	1	1
> KI1 - Kiribati	2	1
> Regional	2	1
> SB1 - Solomon Islands	1	1
> Various	1	1
> VU1 - Vanuatu	3	1
<b>South &amp; West Asia Division (SWD)</b>	<b>13</b>	<b>7</b>
> ACP - Anti-Corruption	1	1
> AFG - Afghanistan	4	2
> BD1 - Bangladesh	3	1
> LK1 - Sri Lanka	2	1
> NP1 - Nepal	1	1
> PK1 - Pakistan	2	1
<b>South-East Asia Mainland &amp; Regional Division (SRD)</b>	<b>17</b>	<b>9</b>
> AP1 - ASEAN	1	1
> Regional	3	1
> KH1 - Cambodia	3	2*
> LA1 - Laos	4	2

\* As noted in the body of this annex, during the assessment process one of these documents was found to be a research piece rather than an operational evaluation. It was subsequently excluded from the sample.

> Mekong	3	1
> MM1 - Burma	1	1
> VN1 - Vietnam	2	1
South-East Asia Maritime Division (SED)	8	4
> ET1 - Timor Leste	1	1
> Indonesia	3	1
> PH1 - Philippines	4	2

# Appendix C: Pro-forma for assessing the quality of operational evaluations

## Cover sheet

1. Reviewer
2. Investment name and number
3. Cluster evaluation (Y/N)
4. Number of evaluation team members
5. Evaluation is partner-led, joint or DFAT-led
6. Evaluation team leader skills
7. Evaluation team skills (other team members)
8. DFAT staff member included on team (Y/N)
9. Evaluation duration:
  - » Fieldwork days
  - » Total person-days
10. Evaluation cost
11. Number of evaluation questions:
12. Performance criteria assessed (Y/N)
  - » Numerical ratings for performance criteria provided (Y/N)
13. Management response located (Y/N)
14. If no management response, are recommendations followed up by other means (eg working groups)
15. Number of recommendations in evaluation report
16. Number of recommendations accepted
17. Number of recommendations partially accepted
18. Number of recommendations not accepted

## Ratings

Satisfactory		Less than satisfactory	
<b>6</b>	Very high quality	<b>3</b>	Less than adequate quality
<b>5</b>	Good quality	<b>2</b>	Poor quality
<b>4</b>	Adequate quality	<b>1</b>	Very poor quality

N/A: The criterion does not apply to the evaluation



Key quality areas and criteria	Quality statements	Rating 1–6	Evidence (include page numbers)
<b>Report features</b>			
1 Executive summary	The executive summary provides all the necessary information to enable primary users to make good quality decisions		
<b>Evaluation purpose and scope</b>			
2 Purpose of evaluation	The purpose of the evaluation is provided, including the overall purpose and primary users of the information		
3 Scope of evaluation	The scope matches the evaluation resources; methods are defined and roles of the team, DFAT management and others are set out.		
Overall comments			
<b>Evaluation methodology</b>			
4 Appropriateness of the methodology and use of sources	The methodology includes justification of the design of the evaluation and the techniques for data collection and analysis. Methods are linked to and appropriate for each evaluation question. Triangulation is sufficient. The sampling strategy is appropriate (where applicable)  Limitations to the methodology and any constraints encountered are described  Ethical issues such as privacy, anonymity and cultural appropriateness are described and addressed		
5 Adequacy and use of M&E	The adequacy of M&E data/systems are described. The evaluation makes use of the existing M&E data.		
Overall comments			
<b>Findings, conclusions and recommendations</b>			
6 The context of the initiative	The context of the initiative is described (including policy, development and institutional context) and its influence on performance is assessed.		
7 Evaluation questions	The report identifies appropriate evaluation questions and then answers them. An appropriate balance is made between operational and strategic issues.		
8 Credibility of evidence and analysis	Findings flow logically from the data, showing a clear line of evidence. Gaps and limitations in the data are clearly explained. Any assumptions are made explicit.  Conclusions, recommendations and lessons are substantiated by findings and analysis. The relative importance of findings is stated clearly. The overall position of the author is unambiguous  In assessing outcomes and impacts, attribution and/or contribution to results are explained. Alternative views / factors are explored to explain the observed results		
9 Recommendations	Conclusions, recommendations and lessons are clear, relevant, targeted and actionable so that the evaluation can be used to achieve its intended learning and accountability objectives.  Any significant resource implications are estimated		

Overall comments	
------------------	--

Good practice example?		(y/n)	Evidence (include page numbers)
	<p>Evaluation report</p> <p>Does the evaluation report represent a good example of evaluation practice, and if so why and in what areas?</p>		
	<p>Evaluation plan</p> <p>If the evaluation plan is available, does it provide a good example of a detailed plan to conduct the evaluation?</p>		
	<p>Terms of reference</p> <p>If the terms of reference are available, do they provide a clear background and rationale, and a list of prioritised evaluation questions?</p>		
	<p>Management response</p> <p>Is the management response fairly concrete and clearly indicate intended actions? Does it clearly state if management agrees or disagrees with the recommendations? Does it indicate by who and when things would be done?</p>		

Evaluation snapshot?		(y/n)	Evidence (include page numbers)
Does the evaluation fulfil one or more of the following criteria:	Uses innovative evaluation methods		
	Credibly demonstrates achievement of higher order outcomes		
	Common theme occurring in a group of operational evaluations focussing on the same sector		
	'Outlier' evaluation from a group of operational evaluations focussed on the same sector (eg examines an innovative investment or has valuable findings not seen in other evaluations)		
	Uncovers unexpected domains of change, e.g. findings that would not normally be reported in regular performance management systems such as AQC's		
	Demonstrate good use of data from investment monitoring and evaluation systems		
	Led to a significant change in the investment or department		
	Other valuable learnings for the department and the broader aid community: please specify:		

## Appendix D: Collection and analysis of investment and evaluation characteristics

Relevant review question	Characteristic	Data source	Analysis  Note: this will include comparisons to the findings of the first Review
<b>Investment characteristics</b>			
1; 7	Investment value	Evaluation database	Range of investment values  Average investment value (including by sector)
1	Investment sector	Evaluation database	Number/percentage of evaluations by sector
1	Investment country/region	Evaluation database	Number/percentage of evaluations by country/region
2	Quality of M&E systems	Aidworks (M&E ratings from Aid Quality Checks completed in 2015)	Relationship between quality of M&E systems and evaluation quality
<b>Evaluation characteristics</b>			
1	Evaluation purpose (eg progress or completion evaluation)	Evaluation database	Number/percentage of evaluations by purpose
1	Partner-led or joint evaluation	Evaluation database	Number/percentage of evaluations that are partner-led, joint, or DFAT-led
1	Cluster evaluation	Evaluation database	Number/percentage of cluster evaluations
1; 2	Evaluation cost	Aidworks. If not available in Aidworks, a cost estimate can be calculated using the advisor remuneration framework.	Range of evaluation cost  Average evaluation cost  Relationship between evaluation cost and evaluation quality
1; 2	Evaluation duration (total time and time in the field)	Evaluation documents (to be recorded in pro-forma)	Range of evaluation duration  Average evaluation duration  Relationship between evaluation duration and evaluation quality

<b>1; 2</b>	Evaluation team size	Evaluation documents (to be recorded in pro-forma)	Range of evaluation team size Average evaluation team size Relationship between evaluation team size and evaluation quality
<b>1; 2</b>	Evaluation team composition <ul style="list-style-type: none"> <li>- Presence of specialist evaluator skills</li> <li>- Presence of DFAT staff on team</li> </ul>	Evaluation documents (to be recorded in pro-forma)	Number/percentage of evaluations which include specialist evaluator skills Number/percentage of evaluations which are led by a specialist evaluator Number/percentage of evaluations with DFAT staff on team Relationship between presence of specialist evaluation skills, and DFAT staff on team, with evaluation quality
<b>1; 2</b>	Number of evaluation questions	Evaluation documents (to be recorded in pro-forma)	Relationship between number of evaluation questions and evaluation quality
<b>1</b>	Aid criteria (relevance, effectiveness etc) assessed in the evaluation <ul style="list-style-type: none"> <li>- Including whether numerical ratings are provided</li> </ul>	Evaluation documents (to be recorded in pro-forma)	Number/percentage of evaluations which assessed and/or rated aid criteria
<b>5</b>	Publication of evaluation report	Evaluation database	Number/percentage of evaluations which have been published on the DFAT website
<b>5</b>	Month evaluation is finalised	Evaluation database	Used to facilitate analysis of differences between evaluations completed before and after the introduction of the new evaluation policy, if required
<b>6</b>	Number of recommendations in evaluation report  Number of recommendations accepted (in management response)	Evaluation documents (to be recorded in pro-forma)	Percentage of recommendations which have been accepted, partially accepted and not accepted

	Number of recommendations partially accepted (from management response)		
	Number of recommendations not accepted (from management response)		

## Appendix F: Factors that contribute to evaluation quality identified in other meta-evaluations

	First DFAT Review of Operational Evaluations (2014)	USAID Meta-evaluation of quality and coverage of USAID evaluations 2009-2012 (2013)	Are SIDA evaluations good enough? (2008)	NORAD Can we demonstrate the difference that Norwegian Aid Makes? (2014)	DFID Rapid Review of Embedding Evaluation (2014)
<b>Summary of meta-evaluation</b>	Examined the quality of 87 operational evaluations completed in 2012.	Examined the quality of 340 evaluations conducted between 2009 – 2012.	Examined the quality of 34 operational evaluation reports published between 2003 and 2005.	Examines why it is difficult to determine the results (ie outcomes and impacts) of Norwegian aid.	A review of DFID's program to 'embed evaluation' within and strengthen the evidence base of its programs.
<b>Factors related to evaluation quality</b>					
<b>1. Duration of an evaluation (total days and field days)*</b>	Longer evaluations (including more fieldwork days) associated with higher evaluation quality (p. 34-5; statistical analysis). Note that reliable cost data was not available.	The meta-evaluation believes that cost and duration influence quality, but did not have the data to test this (p. 27).	The meta-evaluation found no clear correlation between evaluation cost and quality, and disputes whether cost is actually a factor that would drive quality (p. 24)		
<b>2. The purpose of an evaluation (e.g. progress or completion evaluation)</b>	Progress reports are higher quality than completion reports (p. 33; statistical analysis)				Quality of evaluation depends on whether it is fit for intended purposes (p. 40; an assumption and no clear evidence to support it is provided)
<b>3. Evaluation team size</b>	The optimal team size for quality evaluation is 3-4 team members (p. 35; statistical analysis)				
<b>4. Evaluation team composition, particularly the presence of specialist evaluator skills</b>	Teams with a strong team leader and expertise include covering evaluation expertise particularly evaluation expertise, leadership, and sectoral knowledge are associated with higher evaluation	The presence of an evaluation specialist on the team is linked to better evaluation quality (p. 25; statistical analysis)	Evaluator competencies are assumed to influence evaluation quality (not tested statistically)		

\* Note evaluation cost is seen as a proxy for evaluation duration, and so has not been included as a standalone factor

	quality (p. 36; qualitative analysis)				
<b>5. Quality of investment M&amp;E systems</b>	The quality of an investment's M&E system is assumed to drive evaluation quality (p. 32; no statistical analysis possible)				
<b>6. The number of evaluation questions addressed</b>		Previous USAID meta-evaluations found a greater number of evaluation questions led to lower quality, but the 2013 found statistical <u>no</u> association of this manner (p. 27)	Large numbers of evaluation questions are assumed to result in lower evaluation quality (not tested statistically; p. 72)		

## Appendix G: Survey on evaluation capability in DFAT

The Office of Development Effectiveness (ODE) is currently conducting a review of DFAT's operational evaluations. This review aims to help the department understand the practices related to, and the quality of, operational evaluations.

This survey is part of the review. It is for internal use, aimed at better understanding the capabilities of DFAT staff to commission and conduct evaluations. Your response will help inform ongoing discussion on evaluation arrangements in DFAT.

Survey responses will be reported in aggregate. Personal identifying information is optional, and will not be used for reporting purposes.

The survey will take around 5 minutes to complete.

ODE will brief the P&Q Network Meeting on the results of the survey in due course.

On behalf of ODE, thank you for your participation.

There are 13 questions in this survey

### Evaluation knowledge and experience

#### 1) Have you been involved in commissioning evaluations?

Commissioning an evaluation includes writing terms of reference, contracting an evaluation team and reviewing their reports and other deliverables. Please choose all that apply.

- No
- I have developed, or provided substantial input to, an evaluation terms of reference
- I have provided substantial input to selecting an evaluation team
- I have contributed to the review of an evaluation plan against the DFAT M&E Standards
- I have contributed to the review of an evaluation report against the DFAT M&E Standards
- Other: Please specify [free text]

#### 2) If yes, what is the total number of evaluation you have been involved in, in this capacity?

Please write your answer here: [free text]

#### 3) Have you participated in an evaluation as an evaluation team member other than team leader)?

Being part of an evaluation team includes: providing input to the evaluation plan; participating in a field mission; collecting and analysing data; and drafting an evaluation report. Please choose all that apply.

- No
- I have accompanied an evaluation team during fieldwork as an observer
- I have contributed to an evaluation plan
- I have contributed to data collection and/or analysis
- I have written sections of an evaluation report
- Other: Please specify [free text]



**4) If yes, what is the total number of evaluations you have been involved in, this capacity?**

Please write your answer here: [free text]

**5) Have you led an evaluation team (in DFAT or elsewhere)?**

An evaluation team leader is someone who has overall responsibility for defining the evaluation approach and methods, collecting and analysing data and drafting the evaluation report. Please choose only one of the following:

- Yes
- No

**6) If yes, what is the total number of evaluations you have been involved in, in this capacity?**

Please write your answer here: [free text]

**7) What evaluation methods and approaches have you previously employed?**

Please choose all that apply.

- Not applicable
- Develop a program logic / theory of change
- Develop a sampling strategy
- Design and use qualitative methods for data collection and analysis
- Design and use quantitative methods for data collection and analysis
- Design and use mixed methods for data collection and analysis
- I have managed input from technical specialists
- Other: please specify [free text]

**8) Knowledge and experience of evaluation standards**

Please choose all that apply.

- I am familiar with the DFAT monitoring and evaluation standards
- I have experience applying the DFAT monitoring and evaluation standards
- I am familiar with the Australasian Evaluation Society or other evaluator code of ethics
- None of the above

**9) Do you have training and/or qualifications in evaluation?**

Include courses run by DFAT and other parties. Please choose all that apply.

- No
- Completed at least one short course (less than 5 days)
- Completed multiple short courses (totalling over 5 days)
- Completed at least one in-depth course and/or multiple short courses (10 days or more)
- Currently studying for or have obtained a university degree in evaluation or similar field
- Other: Please specify [free text]

## Other information

### 10) Optional: Your name

Please write your name here:

### 11) What is your substantive level?

- € APS 5-6
- € EL1
- € EL2
- € SES
- € LES (please type position level in comment box)

Make a comment on your choice here: [free text]

### 12) How long have you worked at DFAT? Please include employment with AusAID.

Please choose only one of the following.

- € Less than 1 year
- € 1-2 years
- € 3-5 years
- € 6-10 years
- € More than 10 years

### 13) In which division do you work?

- |        |         |
|--------|---------|
| € ACD  | € MAD   |
| € AMD  | € MDD   |
| € APO  | € MPD   |
| € ASNO | € NAD   |
| € AUB  | € ODE   |
| € CCD  | € OPO   |
| € CMD  | € OTN   |
| € DPD  | € PAD   |
| € EUD  | € PCD   |
| € EXB  | € TED   |
| € FTD  | € SED   |
| € HMD  | € SRD   |
| € IMD  | € SWD   |
| € ISD  | € Other |
| € LGD  |         |

## Annex B: Operational evaluations reviewed

Investment number	Evaluation title	Country	Primary sector	Evidence and analysis assessed as credible?	Published on DFAT website?
ING310	Laos Australia NGO Cooperation Agreement (LANGOCA) Program	Laos	Multi-sector/Cross-Cutting	Yes	Yes
INH157	ASEAN-Australia Development Cooperation Program Phase II	Multi-country	Trade Policy and Regulations and Trade-Related Adjustment	No	Yes
INH329	Enterprise Challenge Fund for Pacific and South East Asia	Multi-country	Multi-sector/Cross-Cutting	Yes	Yes
INH946	Independent Completion Review of the Philippines Response to Indigenous Peoples' and Muslim Education (PRIME) Program	Philippines	Education	Yes	Yes
INI486a	Evaluation of the ICRC Special Fund for the Disabled	Multi-country	Government and Civil Society	Yes	No
INI486b	World Health Organisation Value for Money Assessment	Multi-country	Government and Civil Society	Yes	No
INI486c	United Nations Development Program Value for Money Assessment	Multi-country	Government and Civil Society	No	No
INI598, INK586, INI194 & INI767	Review of Australian aid initiatives in the Pacific aimed at ending violence against women	Multi-country	Government and Civil Society	Yes	Yes
INI620	Kiribati Education Improvement Program	Kiribati	Education	Yes	Yes
INI820	Evaluation of the Pacific Sub Regional Programme on Violence against Women and Girls 2011-2014	Multi-country	Government and Civil Society	Yes	No
INI864	Reducing Rheumatic Fever and controlling Rheumatic Heart Disease in four Pacific Island nations	Multi-country	Government and Civil Society	Yes	No
INI946	Mid-Term Review of the MRC Programmes Climate Change and Adaptation	Multi-country	Multi-sector/Cross-Cutting	No	Yes
INI952	Improved food and nutrition security of vulnerable households in Zimbabwe through market-based input assistance FAO Agricultural Inputs Provision Programme 2012-14 Project Completion Review	Zimbabwe	Agriculture	No	Yes

Investment number	Evaluation title	Country	Primary sector	Evidence and analysis assessed as credible?	Published on DFAT website?
IN1953, INJ860, INJ826, ING310	Independent Evaluation of Australia's Support to the UXO Sector in Laos	Laos	Government and Civil Society	Yes	Exempted from publication
INJ129	Clinton Health Access Initiative (CHAI) Vietnam: Phase III Mid-Term Review	Vietnam	Population Policies/Programmes and Reproductive Health	Yes	Yes
INJ137	Independent Progress Review of DFAT Law and Justice Assistance in Indonesia	Indonesia	Government and Civil Society	Yes	Yes
INJ318	Australia-Africa Community Engagement Scheme Mid-Term Review	Pan Africa	Multi-sector/Cross-Cutting	No	Yes
INJ346	Formative Evaluation Report Joint Engagement between WFP and AusAID in Cox's Bazar, Bangladesh	Bangladesh	Humanitarian Aid	No	No
INJ485	Joint Review of the Solomon Islands National Transport Fund	Solomon Islands	Transport and Storage	Yes	Yes
INJ636	Final evaluation TI Asia Pacific Regional Programme	Multi-country	Government and Civil Society	Yes	No
INJ640	High level Strategic Review Fiji Health Sector Support Program	Fiji	Health	Yes	No
INJ768	MAEPA Development Assistance Facility for Afghanistan (DAFA) DAFA I, II, III	Afghanistan	Multi-sector/Cross-Cutting	Yes	Exempted from publication
INJ785	Independent evaluation for early childhood care and education in Khyber Pakhtunkhwa (ECCE-KP)	Pakistan	Education	Yes	Yes
INJ857	Children of Uruzgan Mid-Term Review	Afghanistan	Multi-sector/Cross-Cutting	Yes	Yes
INJ889	Nepal Public Financial Management Support Multi Donor Trust Fund (MDTF) Mid-term evaluation	Nepal	Government and Civil Society	No	Yes
INJ898	Rapid Appraisal Report for the National Library and Archive Building	Vanuatu	Other Social Infrastructure and Services	Yes	No

Investment number	Evaluation title	Country	Primary sector	Evidence and analysis assessed as credible?	Published on DFAT website?
INK078	Mid-Term Review of the Philippines-Australia Public Financial Management Program (PFMP)	Philippines	Government and Civil Society	Yes	Yes
INK103	Sri Lanka Community Forestry Project Mid-Term Review	Sri Lanka	Forestry	No	Yes
INK178	Mid Term Review of the DFAT - TAF Partnership	Multi-country	Multi-sector/Cross-Cutting	Yes	Exempted from publication
INK211	Roads for Development (R4D) Program Final Mid-Term Review	Timor-Leste	Transport and Storage	Yes	Yes
INK235	Mid-Term Review of International Mining for Development Centre (IM4DC)	Multi-country	Mineral Resources and Mining	Yes	Yes
INK437	Improving Health Service Delivery in Myanmar: UN Joint Program on Maternal, Newborn and Child Health	Burma	Health	Yes	Yes
INK509, INL291	Australia's humanitarian response to the Syria Crisis	The Middle East	Humanitarian Aid	Yes	Yes
INK523b	Partner-led Mid-Term Review of Sweden's Support to Transparency International Cambodia	Cambodia	Government and Civil Society	Yes	Yes
INK733	Independent Progress Review of the Regional HIV Capacity Building Program	Multi-country	Government and Civil Society	Yes	No