



Lessons from the Review of Operational Evaluations

What are operational evaluations?

Operational evaluations focus on individual aid activities. They are commissioned and managed by DFAT's country or regional programs. Operational evaluations must be independent, meaning they must be conducted by a team which is external to the commissioning program.

In the Department of Foreign Affairs and Trade, independent evaluations have taken place at two levels:

- » Strategic evaluations are produced by the Office of Development Effectiveness (ODE). These are high-level evaluations of aid program policies, strategies and approaches to common development issues
- » Operational evaluations are managed by country and regional programs. These focus on individual aid investments.

The Review of Operational Evaluations examined 35 operational evaluations completed in 2014. This was the second Review of Operational Evaluations; the first Review examined 87 operational evaluations completed in 2012.

The Review of Operational Evaluations completed in 2014 had three objectives:

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

Findings

The Review had five main findings.

Operational evaluations use modest financial and consultant resources

The first finding was that operational evaluations use modest financial and consultant resources. We found that the average cost of operational evaluations completed in 2014 was \$80,000. As shown in Chart 1, this represents a decrease from 2012. Operational evaluations also cost, on average, a modest 0.37% of investment value.

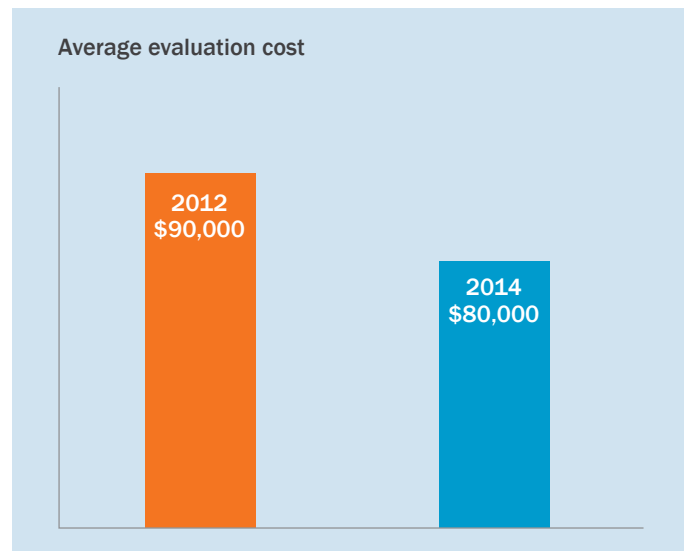


Chart 1: The average cost of operational evaluations decreased between 2012 and 2014

On average, evaluation teams worked on each operational evaluation for a total of 72 working days (for example, two team members may have worked for 36 days each). This includes an average of 32.5 fieldwork days.

The proportion of operational evaluations conducted jointly by DFAT and one of our partners, or led by one of our partners, also increased from 17% in 2012 to 34% in 2014. In addition, the proportion of cluster evaluations, or evaluations which examine more than one investment, increased from 7% in 2012 to 14% in 2014. Both cluster and partner-led/joint evaluations allow for evaluation costs to be shared across investments or between partners, reducing the financial resources DFAT needs to commit to operational evaluations.

For seven of the nine quality criteria, we found that at least 70% of evaluations had adequate or better quality.

The quality of operational evaluations is good

The second finding was that the quality of operational evaluation is good. We assessed the quality of operational evaluations against nine quality criteria, such as clarity of evaluation purpose, appropriateness of methodology, and the relevance and clarity of recommendations.

For seven of the nine quality criteria, we found that at least 70% of evaluations had adequate or better quality. We also found that, compared to 2012, evaluation quality improved or stayed steady against six of the nine criteria.

We also identified one criteria as the proxy for overall evaluation quality. This criteria was ‘credibility of evidence and analysis’, which examined whether findings flow logically from the data. 77% of operational evaluations were assessed as adequate or better quality against this criteria. This demonstrates the quality of evaluations is generally good, but that there has been no significant improvement from 2012 when 74% of operational evaluations were also assessed as being of adequate or better quality (see Chart 2).

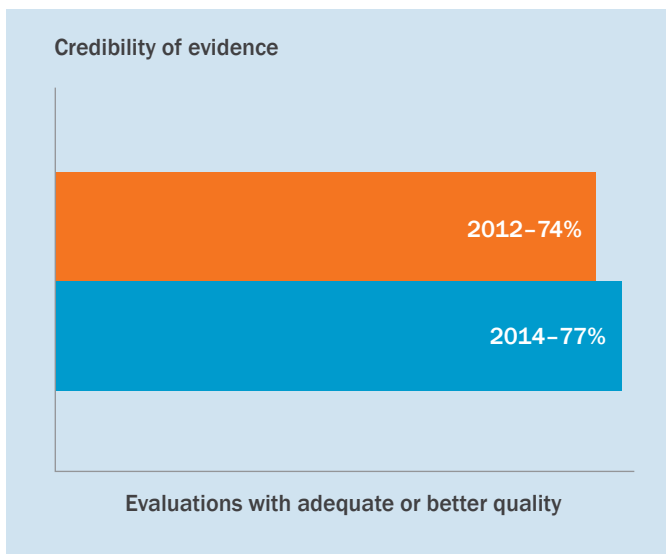


Chart 2: The quality of operational evaluations has remained steady

A further finding was that operational evaluations commissioned and managed by DFAT were more likely to be higher quality than evaluations conducted jointly by DFAT and one of our partners, or led by one of our partners. Although such joint and partner-led evaluations may have lower costs (as highlighted in the first finding), their lower quality raises questions about the value they provide for DFAT.

For evaluation quality, the key is ‘everything in moderation’

The third finding was that for evaluation quality, the key is ‘everything in moderation’. We found that evaluations which had a moderate duration, number of evaluation questions and team size were more likely to be adequate quality than those evaluations which were very large. Evaluations which were poorly resourced (for example, with a very short duration or a very small team) were also much more likely to be poor quality.

Three other factors also influenced evaluation quality:

- » Evaluations undertaken at the end of an investment (known as completion evaluations) were more likely to be adequate quality than evaluations undertaken during an investment (known as progress evaluations)
- » When monitoring and evaluation expertise was included on an evaluation team, the evaluation was more likely to be adequate quality, and
- » As shown in Chart 3, when a DFAT staff member was included on the evaluation team the evaluation was more likely to be adequate quality*.

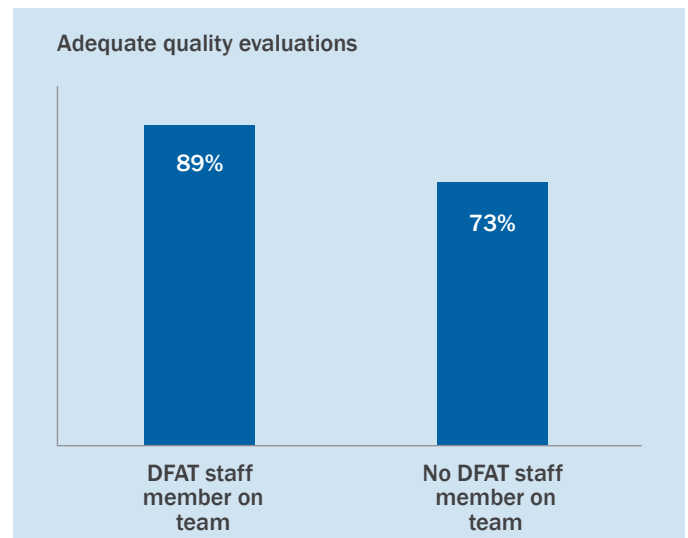


Chart 3: Including DFAT staff on evaluation teams may improve evaluation quality

There are questions about how well operational evaluations are used

The fourth finding was that there are questions about how well operational evaluations are used. One issue relates to whether the findings of operational evaluations are used to improve investments. Each operational evaluation should have a management response which outlines whether DFAT agrees with the recommendations and how they will be implemented. However, we were only able to locate management responses for half of the evaluations examined in the Review, indicating evaluation findings are not being followed up and used.

* Note the Review only included a small number of evaluations which included a DFAT staff member, so this finding should be treated with some caution.

We also found that the evaluation recommendations were one of the lowest quality aspects of operational evaluations. Only 66% of evaluations were found to have recommendations that were clear, relevant and actionable, meaning the remaining one third of evaluations did not have adequate quality recommendations. This may be influencing the low number of management responses.

In addition, it is DFAT policy that all evaluations should be published on the DFAT website. However, by July 2015 only 38% of operational evaluations completed in 2014 had been published. As shown in Chart 4, this is consistent with the low publication rate identified in the previous Review.

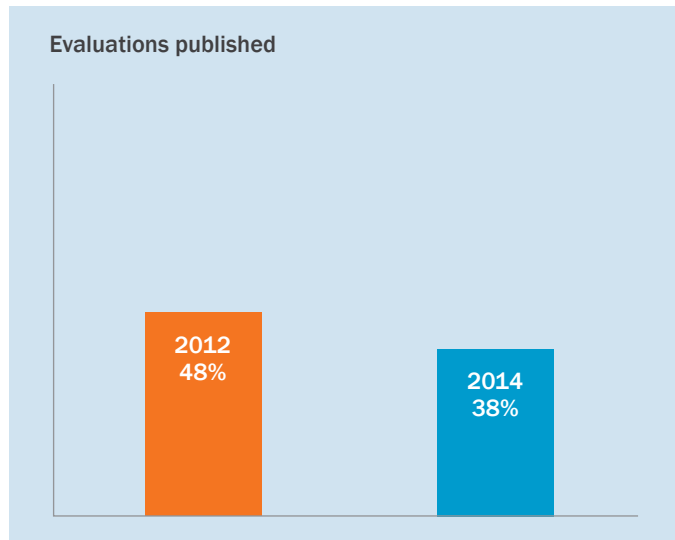


Chart 4: The publication rate for operational evaluations remains consistently low**

DFAT does not have sufficient evaluation capacity to allow large numbers of evaluations to be conducted in-house

The final finding was that DFAT does not have sufficient capacity to allow a large number of evaluations to be conducted in-house.

We conducted an all-staff survey to assess DFAT's capability to commission and conduct evaluations. The survey found that DFAT has some evaluation capability. A number of staff have experience commissioning evaluations and applying DFAT's evaluation quality standards, and have received training in evaluation.

However, a much smaller number of respondents have been a member of an evaluation team or have led an evaluation. Staff with experience leading evaluation teams are also widely dispersed across DFAT. As a result, it appears likely DFAT will continue to need to use consultants as members of evaluation teams.

** The publication rate of 48% for evaluations completed in 2012 was measured in September 2013, nine months after the calendar year ended. The publication rate of 38% for evaluations completed in 2014 was measured in July 2015, seven months after the calendar year ended. This may explain why the 2012 publication rate is slightly higher—because programs had an additional two months to publish their evaluations.

Tips for maximising the quality of operational evaluations



Evaluation Scope

Keep the scope of the evaluation contained.



Aim for a moderate number of evaluation questions. Evaluations with **11-14 questions** (including sub-questions) tend to be higher quality.



Aim for a moderate evaluation duration. Evaluations with around **60 working days** (in total for all team members) are generally better quality.



Evaluation Team

Think about the size and skill set of your evaluation team.



Aim for a moderately sized evaluation team. Teams of **2-3 people** generally produce better quality evaluations.



Include **monitoring and evaluation** expertise on the evaluation team, as this improves evaluation quality.



Evaluation teams which include **DFAT staff** members may also be higher quality.

Recommendations

It is clear from the Review's findings that operational evaluations are not being used to their full potential and that more needs to be done to ensure they are used to improve DFAT's work. Better oversight mechanisms are also needed to address the low number of management responses and the poor publication rates.

To improve the use of operational evaluations, the Review recommends that:

1. DFAT's Executive has the opportunity to consider planned, completed and published evaluations. This should be provided through regular reporting from the Office of Development Effectiveness.
2. Senior managers (particularly SES Band 1) have direct involvement in deciding what is evaluated in their program, to ensure relevant and useful evaluations are conducted.

The Review also highlights that in the long term DFAT should aim to improve the evaluation capacity of its staff. This will reduce reliance on consultants and partner-led or joint evaluation processes. It will also improve the quality and use of evaluations by DFAT, particularly the quality of recommendations. To address these points it is recommended that:

3. DFAT staff are included on evaluation teams to the extent possible
4. An updated evaluation policy, and associated guidance and reporting, focuses on encouraging recommendations which are clear, relevant and feasible.

The Review made these four recommendations to DFAT's Independent Evaluation Committee. They have endorsed these recommendations and asked ODE to take them into account as part of ODE's role in managing DFAT's evaluation policy.

Learning from Operational Evaluations

What are Evaluation Snapshots?

Evaluation snapshots highlight novel or useful lessons on aid and development. These lessons have been taken from the operational evaluations examined as part of the Review.

In addition to examining the quality of operational evaluations, the Review also identified useful or novel lessons on aid and development from the operational evaluations examined. These lessons are summarised in Part 2 of the Review, titled 'Evaluation Snapshots'. We encourage interested parties to examine both the snapshot summaries and the full evaluations to identify useful lessons which are relevant to them.

The snapshots are:

- » The Laos Australia NGO Cooperation Program demonstrates novel approaches to achieving gender equality and highlights the challenges of working with the poorest members of communities
- » The Enterprise Challenge Fund (ECF) for the Pacific and South East Asia outlines lessons on working with the private sector and the cost-effectiveness of ECFs, particularly in the Pacific
- » The challenges and constraints to cascade training for teachers are highlighted in the Malaysia Australia Education Project for Afghanistan
- » Three operational evaluations provide different approaches to measuring value for money, and
- » The Regional HIV Capacity Building Program demonstrates the risks to program quality when program costs are minimised.

The Office of Development Effectiveness (ODE) is an independent branch within the Department of Foreign Affairs and Trade (DFAT). ODE monitors the Australian aid program's performance, evaluates its impact, and contributes to international evidence and debate about aid and development effectiveness. ODE's work is overseen by the Independent Evaluation Committee (IEC), an advisory body that provides independent expert advice on ODE's evaluation strategy, work plan, analysis and reports.

The full evaluation report and DFAT management response can be accessed at www.ode.dfat.gov.au.