



Australian Government
Department of Foreign Affairs and Trade

Office of Development Effectiveness

Review of Operational Evaluations completed in 2014
Part 1, Quality Review

www.ode.dfat.gov.au

Foreword

I am pleased to present to you the Review of Operational Evaluations completed in 2014, which examines the quality of independent evaluations commissioned by program areas in the Department of Foreign Affairs and Trade (DFAT). This is the second such Review; the first examined independent operational evaluations completed in 2012. This examination of the changes in operational evaluations over time makes this Review one of the strengths of DFAT's performance management system.

Between 2012 and 2014, the Australian aid program has undergone significant changes. Most notable have been the integration of the former Australian Agency for International Development (AusAID) into DFAT; the shift in the Australian Government's aid policy priorities; and the changes to the aid budget.

Despite such large scale changes, I am happy to say the quality of DFAT's operational evaluations remains good. They continue to be a credible source of evidence for the aid program while using modest consultant and financial resources. This is a very positive finding and, in my view, is a testament to the competence of DFAT staff, reflecting their commitment to ensuring evaluation plays a key role in program accountability and quality.

However, more remains to be done. I believe the poor completion of management responses and the low publication rate for evaluations must be addressed with urgency. DFAT's evaluation capacity should also be improved to reduce reliance on external consultants and to increase the use of evaluations.

The Review makes four recommendations to further improve the use and quality of operational evaluations. The Independent Evaluation Committee (IEC), which I chair, endorses these recommendations. They flow logically from the findings of the Review and represent sensible, achievable actions which will improve operational evaluations. I urge the Office of Development Effectiveness and senior policy makers in DFAT act on these recommendations when revising DFAT's evaluation policy.



Jim Adams
Chair, Independent Evaluation Committee

Executive summary

In the Department of Foreign Affairs and Trade (DFAT), independent operational evaluations focus on individual aid investments.

- These evaluations are managed by country and regional programs.

This Review of Operational Evaluations examines 35 such evaluations which were completed in 2014. The Review aims to:

- Provide a better understanding of evaluation practice and quality across DFAT, and
- Facilitate opportunities for learning from operational evaluations.

This is the second Review of Operational Evaluations.

- The first examined 87 operational evaluations completed in 2012.

The Review was a modest exercise undertaken in-house by the Office of Development Effectiveness (ODE). Methods included:

- A desk study to assess the quality of operational evaluations against nine criteria, and to identify good practice evaluations and useful lessons on aid and development
- A survey of DFAT staff to examine the department's evaluation capability
- An audit of evaluation publication conducted by DFAT's Internal Audit Branch.

ODE both conducted this Review and will need to respond to its findings. It is difficult for ODE to draft recommendations which ODE then has to respond to.

- To address this, the Review includes recommendations for endorsement by the Independent Evaluation Committee (IEC).
- If the IEC endorses the Review's recommendations, ODE will act on these when revising DFAT's evaluation policy.

The Review report has two main parts:

- Part 1, Quality Review, discusses the quality of operational evaluations and the implications for DFAT's evaluation policy
- Part 2, Evaluation Snapshots, outlines lessons on aid and development from operational evaluations.

Executive summary: Findings from Part 1, Quality Review

There are five main findings from the Quality Review.

The first is that operational evaluations use modest financial and consultant resources:

- The average cost of an operational evaluation decreased from \$90,000 in 2012 to \$80,000 in 2014
- Operational evaluations cost, on average, just 0.37% of investment value
- The average length of an operational evaluation was only 72 days, including 32.5 fieldwork days.

The second finding is that, although operational evaluations use modest resources, their quality is good:

- Using an overall indicator of evaluation quality, 77% of evaluations were assessed as adequate or better quality
- This demonstrates that the quality of evaluations is generally good, but that there has been no significant improvement from 2012 when 74% of operational evaluations were assessed as adequate or better quality
- It is noteworthy that evaluations managed by DFAT were more likely to be higher quality than evaluations led by a development partner or conducted jointly by DFAT and a development partners.

The third finding is that, for evaluation quality, the key is ‘everything in moderation’:

- Analysis found that evaluations were more likely to be better quality if they had a moderate duration, a modest number of evaluation questions, a medium team size, and monitoring and evaluation expertise on the team
- Evaluation teams which included DFAT staff members were also more likely to be better quality than evaluation teams which did not include DFAT staff
- In addition, it is clear that poorly resourced evaluations are much more likely to be low quality.

The fourth finding is that there are questions about how well operational evaluations are used in DFAT:

- Only 16 of a possible 35 management responses could be located
- By July 2015, only 38% of operational evaluations completed in 2014 had been published on the DFAT website
- Evaluation recommendations were found to be one of the lowest quality aspects of evaluations; this may be influencing the low number of management responses.

Executive summary: Findings from Part 1, Quality Review (continued)

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

- The staff survey shows DFAT has some capacity, with staff participating in training and having experience commissioning evaluations
- However, only a small number have experience leading or participating in an evaluation team, and these staff are widely dispersed across the department.

As ODE was finalising this Review, DFAT decided that the operational evaluations policy should be revised to ensure the time and resources required to produce evaluations is commensurate with the benefits DFAT receives from them.

- The findings of the Review point to areas the revised policy should address.

It is clear that action is needed to ensure operational evaluations are used to their full potential. It is also clear that more robust mechanisms are needed to improve the low rates for completing management responses and publishing evaluations. To address these points it is recommended that:

- 1) DFAT's Executive has the opportunity to consider planned, completed and published evaluations. This should be provided through regular reporting from ODE
- 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.

Further, the Review 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) The updated evaluation policy, and associated guidance and reporting, focuses on encouraging recommendations which are clear, relevant and feasible.

Executive Summary: Findings from Part 2, Evaluation Snapshots

The Review identified five useful or novel lessons on aid and development from the operational evaluations examined.

- These lessons are available to DFAT staff and the broader development community in Part 2, Evaluation Snapshots.

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 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.

Part 1, Quality Review

Contents

Background and methods

Key findings

- Operational evaluations use modest resources
- The quality of operational evaluations is good
- For evaluation quality, the key is ‘everything in moderation’
- There are questions about how well operational evaluations are used
- DFAT does not have sufficient evaluation capacity to allow large numbers of evaluations to be conducted in-house

Recommendations

Evaluation snapshots and annexes

The Review of Operational Evaluations aims to improve our understanding of the practices and quality of operational evaluations

In the Department of Foreign Affairs and Trade (DFAT), independent evaluations are undertaken 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.

This Review focuses on independent operational evaluations. It has 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.

This Review builds on the first Review of Operational Evaluations conducted by ODE

ODE's first Review of Operational Evaluations examined 87 operational evaluations completed in 2012

- These were managed by the Australian Agency for International Development (AusAID).

Since the first Review, the Australian aid program has undergone significant changes:

- AusAID was integrated into DFAT and the policy parameters for operational evaluations subsequently changed in mid-2014 (see table on next slide).
- The primary objective of the policy changes was to ensure DFAT produced fewer evaluations which were more fit-for-purpose.

The aid evaluation policy changed significantly between 2012 and 2014

The changes are summarised in the table below:

Aid evaluation policy	December 2012	July 2014
Monitored initiatives	All investments over \$3m to be independently evaluated at least once in their lifetime.	All investments over \$10m to be independently evaluated at least once in their lifetime.
Criteria	All evaluations to assess investments against six criteria: (i) relevance (ii) effectiveness (iii) efficiency (iv) sustainability (v) impact (vi) gender equality	No requirement to assess investments against particular criteria. Staff have flexibility to focus operational evaluations on the highest priority issues. Note the six criteria are still assessed as part of Aid Quality Checks (AQC)*. In practice, many of the criteria are also covered in operational evaluations as they provide highly relevant information for investment managers.
Ratings	All evaluations should provide a numerical rating between 1-6 for the above criteria (1=very low quality, 6=very high quality).	No requirement to provide numerical ratings. Note that numerical ratings for these criteria, and a series of sub-criteria, are still provided in Aid Quality Checks.

* Aid Quality Checks (AQC) are self-assessments which examine how well investments are performing during implementation. They are conducted on an annual basis for all investments valued at \$3m or more. Each investment is given a rating between 1 (very poor) and 6 (very good) for seven quality criteria: relevance, effectiveness, efficiency, monitoring and evaluation, sustainability, gender equality, and risk management and safeguards. AQC are not publically available.

The current Review was conducted in-house and has a modest scope

There are two main differences between the current Review of Operational Evaluations and the previous Review:

- The team who conducted the Review, and
- The scope of the Review.

ODE conducted the current Review in-house using ODE staff:

- The previous Review used contracted consultants to undertake the work
- ODE conducted the Review in-house to ensure the findings would be relevant to DFAT; any proposed follow-up actions would be appropriate and feasible; and ODE staff would have a strong understanding of current evaluation practice in the department.

To ensure ODE staff could conduct the Review in an efficient manner, it has a modest scope:

- The current Review examines 35 operational evaluations completed in 2014. This is a purposeful sample from the 77 operational evaluations completed in 2014
- The purposeful sample includes at least one operational evaluation from each aid sector, each DFAT aid division and each DFAT aid country program. For more information on the sampling strategy please see Annex A
- This is in contrast to the previous Review which examined all 87 operational evaluations completed in 2012, rather than using a sampling strategy.

The main audiences for the Review are ODE and Performance and Quality staff

ODE will use the Review's findings to improve DFAT's operational evaluations policy, which we oversee.

- ODE and Performance and Quality staff in country programs will also use the Review's findings to provide improved advice to staff on managing high quality operational evaluation processes.

A secondary audience is DFAT's Aid Management and Performance Branch (MPB), which oversees DFAT's investment quality reporting system (of which operational evaluations are part).

The Review examines seven key evaluation questions

The evaluation questions, and the methods used to address them, are summarised in the table below

- Consistent with the modest intent of the Review, the methods used were limited. The main method was a desk review of completed operational evaluations. We chose not to collect significant qualitative data on evaluation process or use.

Evaluation question	Data collection methods
Priority questions	
1. What are the characteristics and quality of operational evaluations? How have these changed since 2012?	<p>Characteristics of evaluations (and related investments) collected from existing ODE evaluation database, DFAT's aid management system (Aidworks), and when assessing each evaluation.</p> <p>Evaluations rated against quality criteria (see next slide).</p> <p>Evaluation characteristics and quality summarised from the first Review of Operational Evaluations completed in 2012.</p>
2. What factors contribute to the quality of operational evaluations?	<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&E systems collected from 2014 Aid Quality Checks.</p>
3. To what degree do operational evaluations provide a credible source of evidence for the effectiveness of the Australian aid program?	Data on evaluation quality collected under Q1 above.
Other evaluation questions	
4. What is the capability to commission and conduct operational evaluations across the department?	Short survey on evaluation capability conducted. The survey was open to all DFAT staff and targeted at the Performance and Quality network.
5. How are the findings from operational evaluations used in the department?	Management responses assessed to examine the proportion of recommendations accepted.
6. How can the publication rate of operational evaluations be improved?	Data on evaluation publication taken from ODE's evaluation tracking tool. Analysis of the DFAT Audit Branch's audit of evaluation publication (2016) conducted.
7. How do DFAT's evaluation policies and practices compare to the evaluation policies and practices of other donors?	Analysis of the OECD Review of Evaluation Systems in Development Cooperation (2016).

To determine evaluation quality, each operational evaluation was assessed against nine quality criteria

The criteria (summarised below) are based on DFAT's Monitoring and Evaluation Standards

- For each criteria an operational evaluation was given a score between 1 (very poor quality) and 6 (very high quality).

Quality criteria	Summary description
1) Executive Summary	The executive summary provides all the necessary information (including on gender issues) to enable primary users to make good quality decisions.
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 of the evaluation matches the evaluation resources. Data collection methods are defined and take into account the needs of groups such as women.
4) Appropriateness of methodology and use of sources	Justification is provided for the data collection and analysis techniques chosen. Consideration is given to analysis of sex-disaggregated data. Triangulation is sufficient and the sampling strategy is appropriate. Limitations to the methods and ethical issues are described and addressed.
5) Adequacy and use of M&E	The adequacy of M&E data and systems are described. Where good quality data is available and relevant to evaluation questions, the evaluation makes use of it.
6) Context of the investment	The context of the investment (including relevant gender issues) is described and its influence on performance is assessed.
7) Evaluation questions	The evaluation 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. Conclusions, recommendations and lessons are substantiated by findings and analysis. Findings relevant to specific groups such as women are included.
9) Recommendations	Conclusions, recommendations and lessons are clear, relevant, targeted and actionable.

There were four main limitations to the Review

These limitations, and how they have been addressed, are summarised in the table below:

Limitation	Management strategy
ODE is assessing the evaluation policy, guidance and support which it oversees and provides.	The Independent Evaluation Committee (IEC) will oversee the Review to help ensure any self-assessment conducted by ODE is defensible.
ODE will need to respond the Review's findings. It will be difficult for ODE to draft recommendations as well as a management response to the recommendations.	The Review report will include recommendations that will be considered by the IEC, rather than by DFAT management. 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.
Review team members need to assess operational evaluations in a consistent manner.	A Review Handbook outlined in detail how assessments should be undertaken. Two moderation meetings were held during the assessment process to ensure team members were assessing evaluations consistently.
The sample size was small, particularly as some data (eg evaluation cost or team size) was not available for all the operational evaluations we assessed.	Some findings and conclusions need to be treated with caution. These are highlighted in this report.

To facilitate learning from operational evaluations, we have identified good practice examples and evaluation snapshots

When assessing operational evaluations, we identified good practice examples which demonstrate the features of high quality operational evaluations:

- The good practice examples are highlighted throughout this report
- They will also be made available on the ODE website.

We also identified operational evaluations which provide useful lessons on aid and development:

- These have been summarised in evaluation snapshots in Part 2 of this report.

In addition, we shared the list of evaluations rated as having adequate or better quality with the Aid Management and Performance Branch and the Development Policy Division in DFAT:

- These parts of DFAT can use such evaluations when assessing the overall quality of the aid program, for example through the Performance of Australian Aid report.

The Review report is divided into three parts

Part 1 (this document) outlines the Review's findings in relation to the quality of operational evaluations, how this has changed since 2012, and how it can be further improved in the future.

Part 2 summarises useful lessons on aid and development in evaluation snapshots.

Part 3 contains the annexes to the report, including a detailed description of the methods used (Annex A) and the list of operational evaluations included in the Review (Annex B).

Contents

Background and methods

Key findings

- Operational evaluations use modest resources
- The quality of operational evaluations is good
- For evaluation quality, the key is ‘everything in moderation’
- There are questions about how well operational evaluations are used
- DFAT does not have sufficient evaluation capacity to allow large numbers of evaluations to be conducted in-house

Recommendations

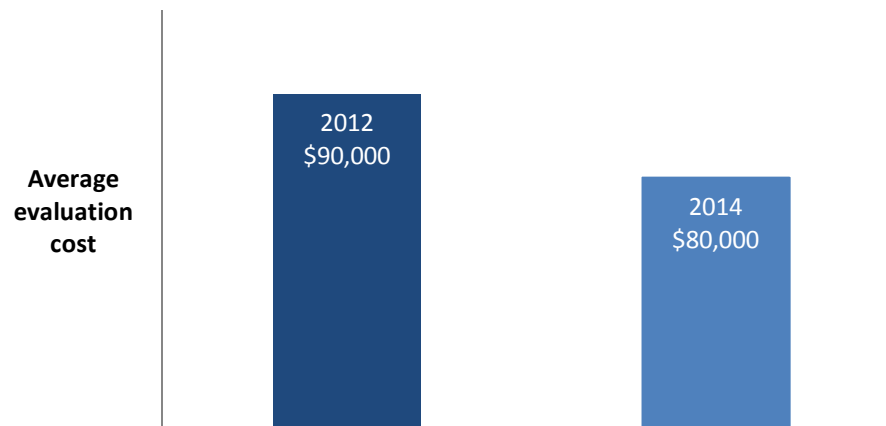
Evaluation snapshots and annexes

The first key finding is that operational evaluations use modest financial and consultant resources (1 of 4)

The median cost of the evaluations we examined was \$53,000 and the average cost was \$80,000. This represents a decrease since 2012, as show in the chart below.

Operational evaluations cost between 0.05% and 0.98% of investment value. The average evaluation cost as a proportion of investment value was 0.37%.

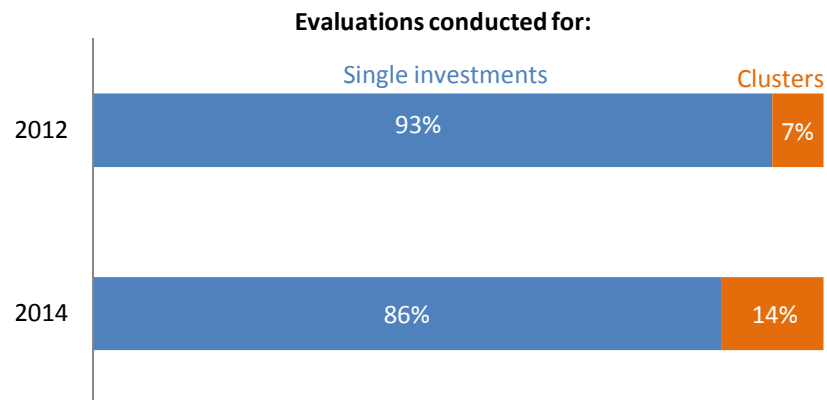
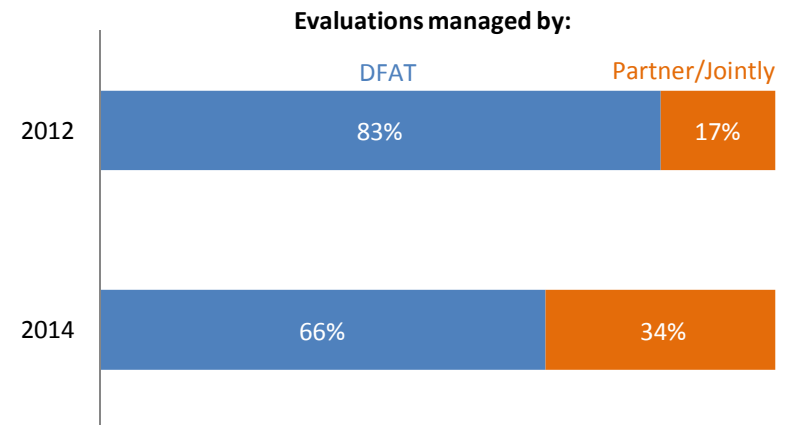
- This is significantly less than the UK's Department for International Development, where evaluations account for a median of 1.9% of investment value (DFID, 2014).



Note: Cost data was available for 25 operational evaluations completed in 2014.

Operational evaluations use modest financial and consultant resources (2 of 4)

The proportion of evaluations led by one of DFAT's partners (for example, another donor), or conducted jointly by DFAT and a development partner, has increased since 2012 (see chart to the right).



The proportion of evaluations which look at a clusters of investments has also increased since 2012 (see chart to the left).

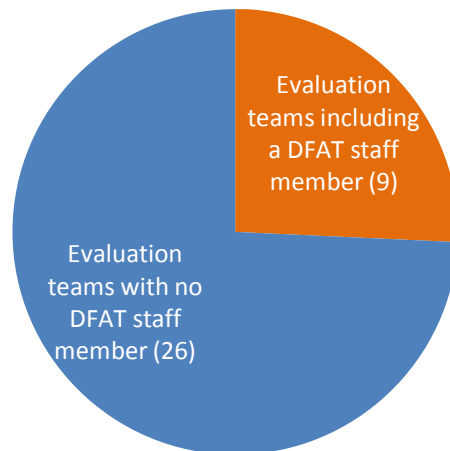
Both cluster and partner-led/joint evaluations allow for evaluation costs to be shared across investments and/or between partners, reducing the financial resources DFAT needs to commit to operational evaluations.

Operational evaluations use modest financial and consultant resources (3 of 4)

On average, the total number of **working days** committed to an operational evaluation by all team members was 72 (for example, two team members may have worked on evaluation for 36 days each). The median was 70 days.

On average, the total number of **fieldwork days** committed to an operational evaluation by all team members was 32.5 (for example, 3 team members may have conducted fieldwork for 11 days each). The median was 29 days.

These days were generally completed by consultants as DFAT staff were only involved in nine of the evaluations, as shown in the chart below:



In five cases, DFAT staff were observers and in the other four cases they played a substantive role in the evaluation team.

Notes: Data on total input days was available for 27 evaluations.

Data on fieldwork days was available for 28 evaluations.

It is not possible to compare results with 2012 because methods for counting evaluation days were not equivalent.

Summary: Operational evaluations use modest financial and consultant resources (4 of 4)

In summary:

- The decreasing cost of operational evaluations
- The small value of operational evaluations as proportion of investment value
- The increasing proportion of cluster and partner-led/joint evaluations, and
- The low average days committed to operational evaluations

lead us to conclude that operational evaluations use modest financial and consultant resources.

Good practice example

The Timor-Leste Roads for Development Program Mid-term Review is a good practice example of a joint operational evaluation. The terms of reference and evaluation plan outline a clear process for managing the evaluation, with an Evaluation Management Committee, made up of representatives from DFAT, the International Labor Organisation and the Government of Timor-Leste, managing the evaluation jointly. As a result, a high quality evaluation report and management response were produced.

Contents

Background and methods

Key findings

- Operational evaluations use modest resources
- The quality of operational evaluations is good
- For evaluation quality, the key is ‘everything in moderation’
- There are questions about how well operational evaluations are used
- DFAT does not have sufficient evaluation capacity to allow large numbers of evaluations to be conducted in-house

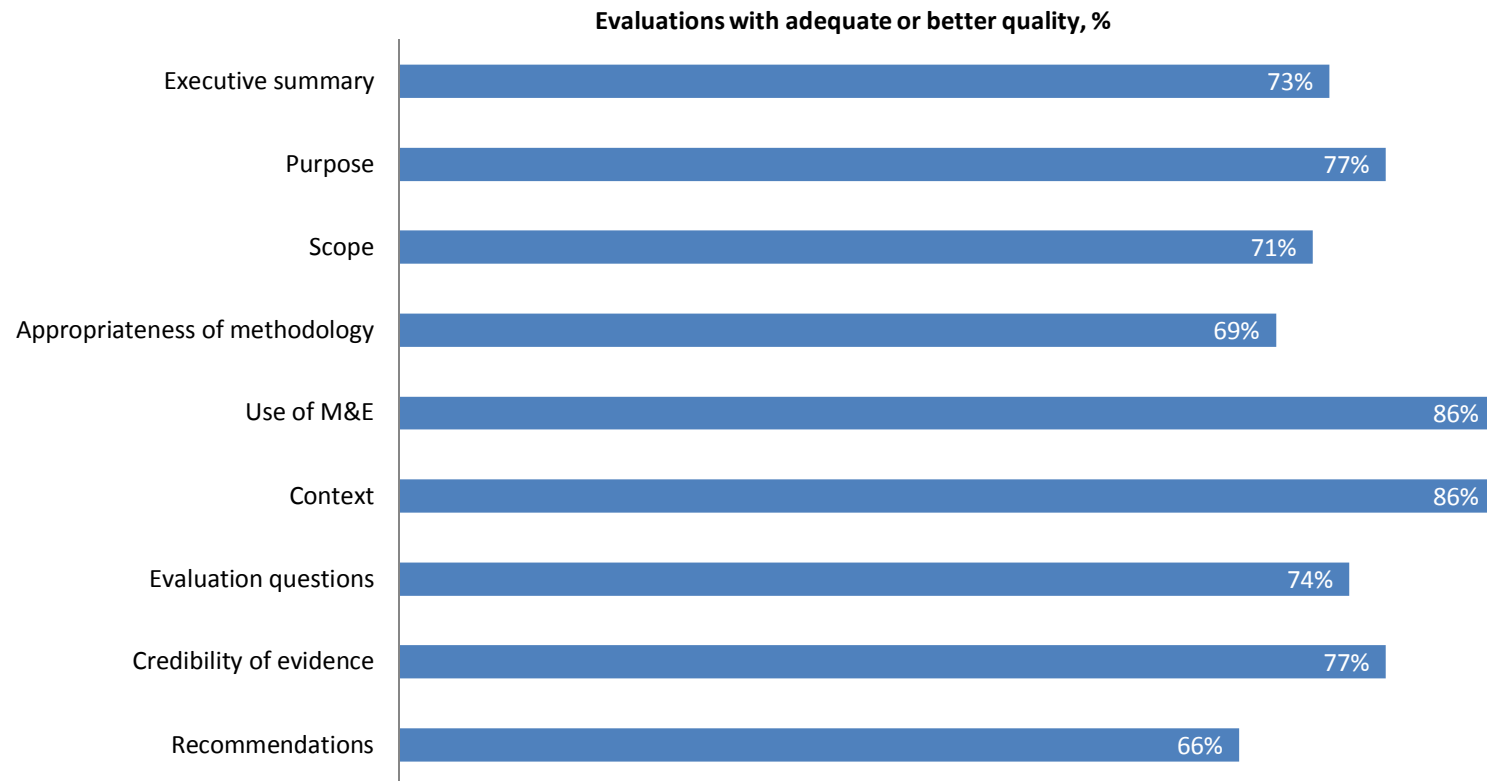
Recommendations

Evaluation snapshots

The second key finding is that the quality of operational evaluations is good and they are a credible source of data for the aid program (1 of 8)

As discussed in the 'Background and methods' section, to determine the quality of operational evaluations we assessed them against nine quality criteria (see slide 11).

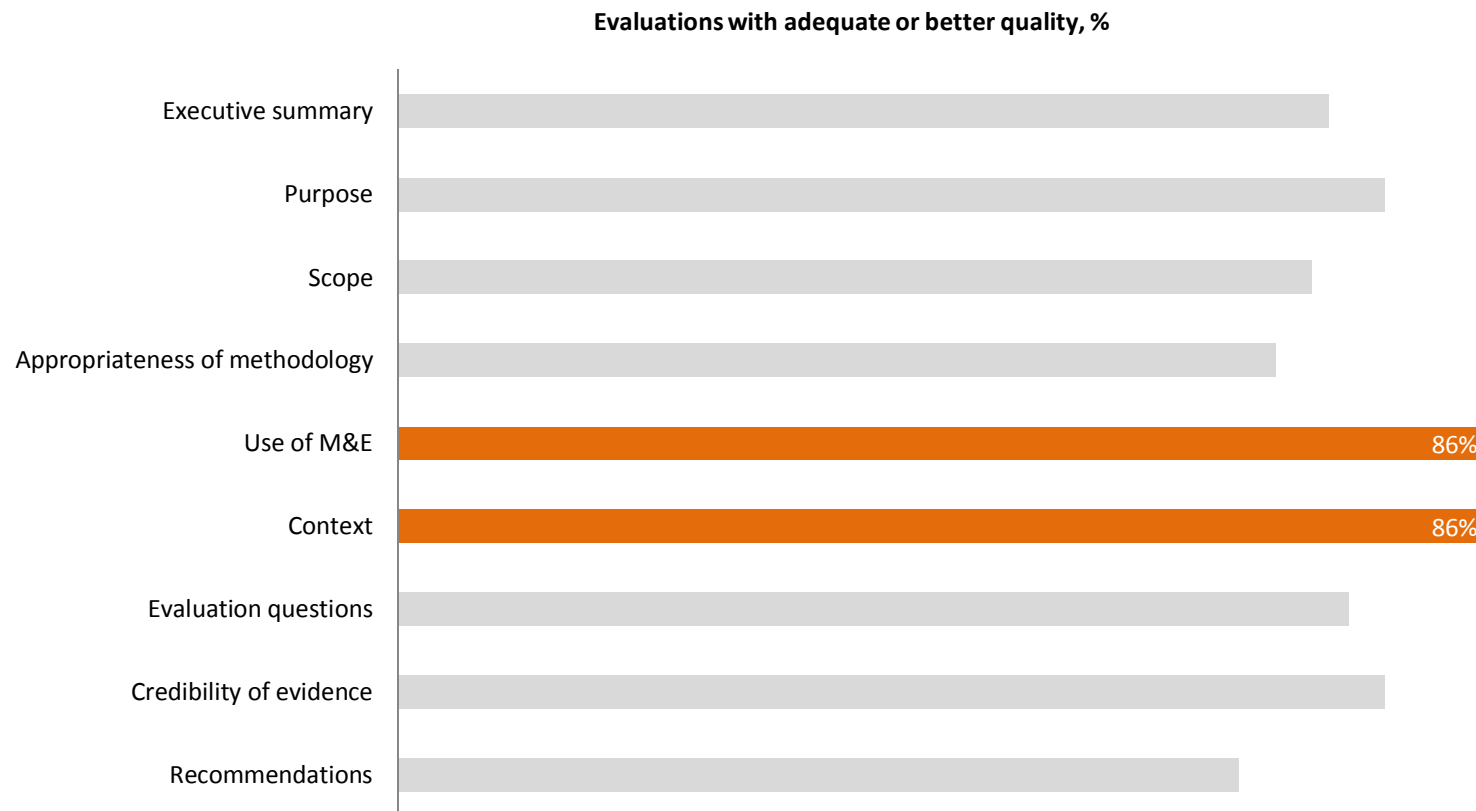
For seven of the nine criteria, at least 70% of evaluations were assessed as adequate or better quality.



The quality of operational evaluations is good and they are a credible source of data for the aid program (2 of 8)

There was a very strong result for two criteria:

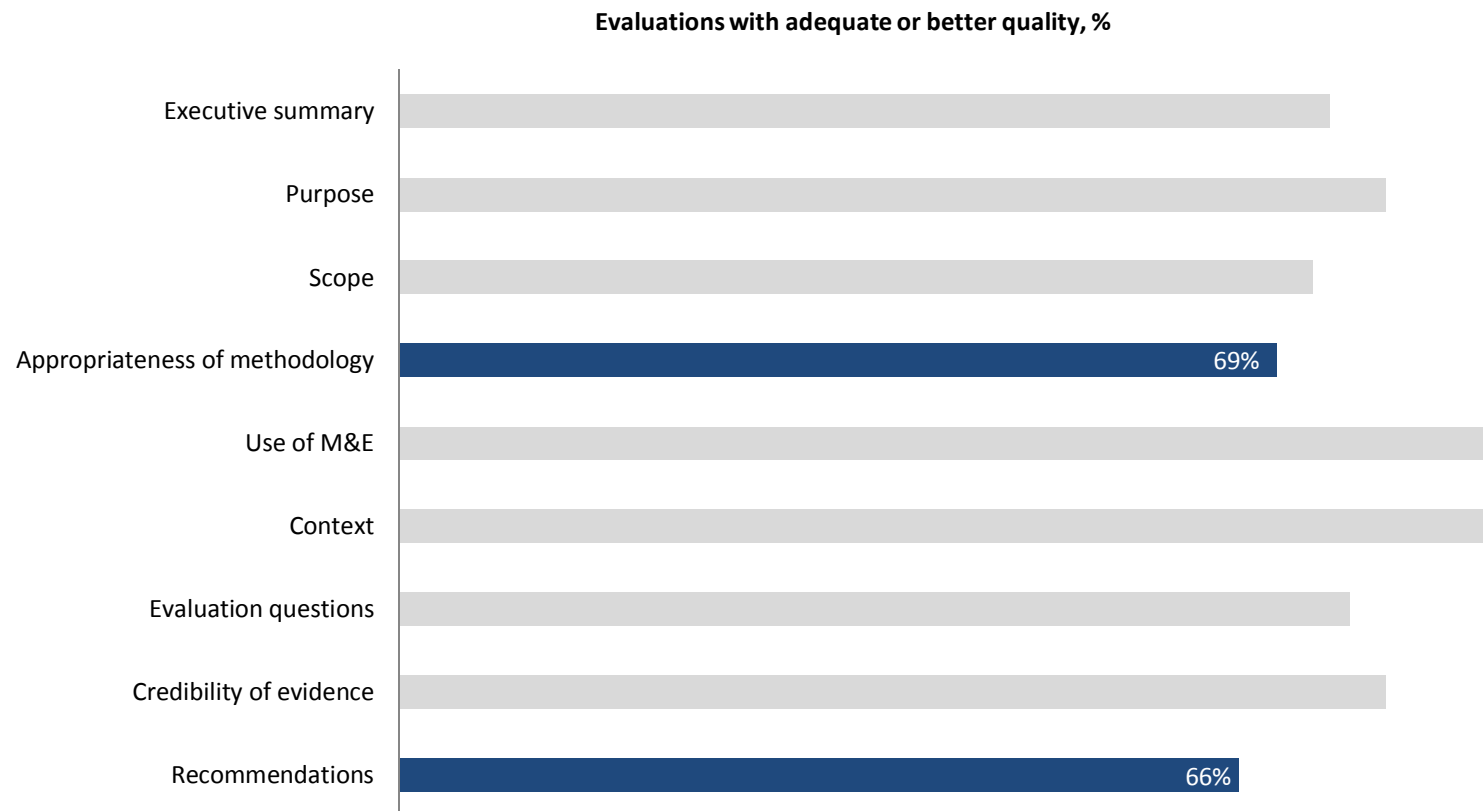
- For ‘use of monitoring and evaluation data’ and ‘context of the investment’, 86% of evaluations were assessed as adequate or better quality.



Note: The criteria ‘use of monitoring and evaluation data’ examined whether an evaluation described the adequacy of an investment’s M&E data and, where good quality data was available and relevant, whether it was used appropriately in the evaluation.

The quality of operational evaluations is good and they are a credible source of data for the aid program (3 of 8)

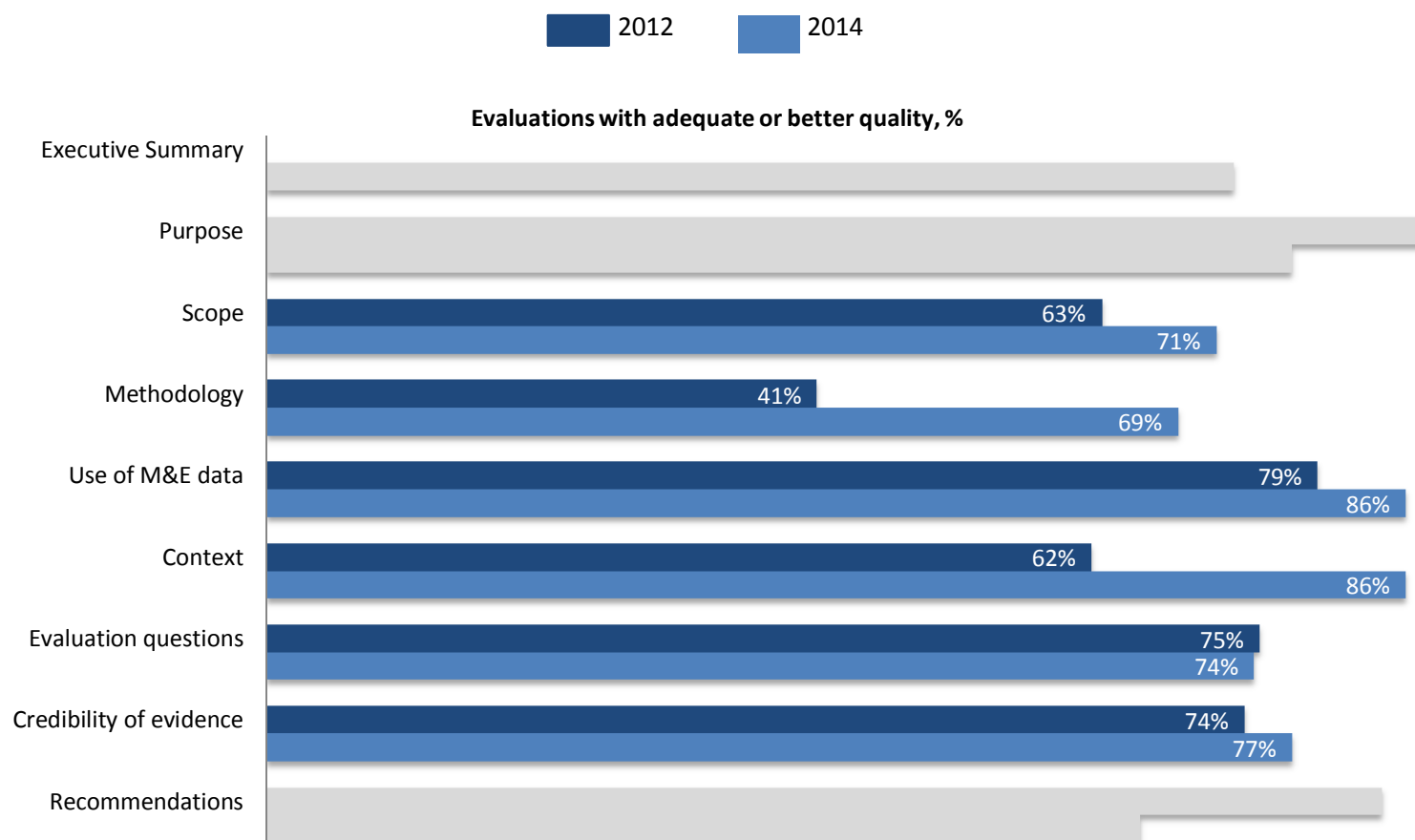
The criteria with the lowest percentage of evaluations with adequate or better quality were 'appropriateness of methodology' (69%) and 'recommendations' (66%).



The quality of operational evaluations is good and they are a credible source of data for the aid program (4 of 8)

Compared to 2012, evaluations assessed as adequate or better quality improved or stayed steady against six criteria, as shown in the chart below.

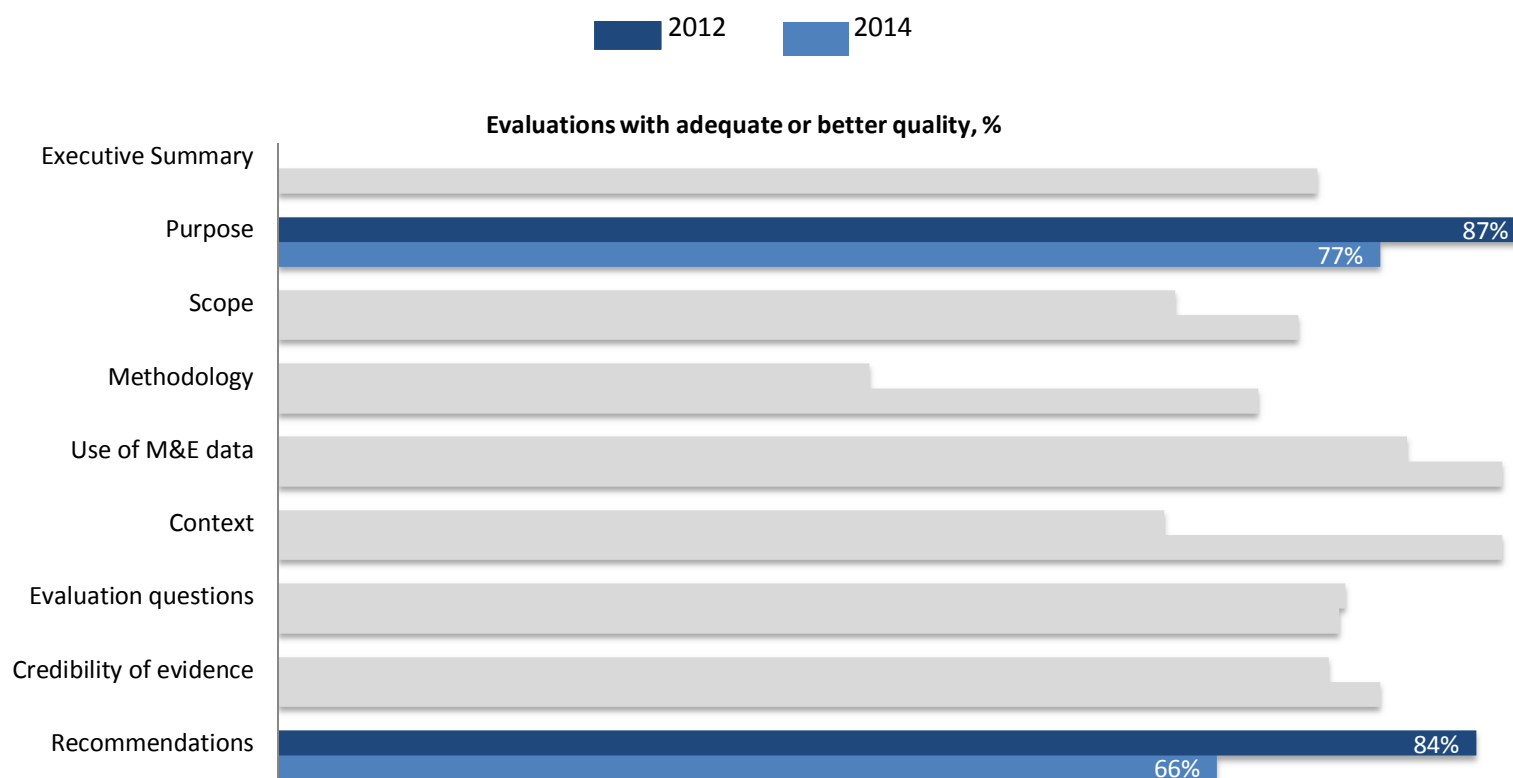
- Although the criteria ‘appropriateness of methodology’ had one of the lowest percentages of adequate quality evaluations, the ratings against this criteria improved significantly since 2012.



The quality of operational evaluations is good and they are a credible source of data for the aid program (5 of 8)

Compared to 2012, evaluations assessed as adequate or better quality declined against two criteria: 'purpose of the evaluation' and 'recommendations'.

- Ratings against the criteria 'recommendations' may have declined not because the quality of recommendations changed, but because the Review team (made up of ODE staff) was better able to assess the relevance and feasibility of recommendations for DFAT, compared to the external consultants who conducted the previous Review.
- Note it is not possible to compare the criterion 'Executive Summary' as this was not assessed in the previous Review.

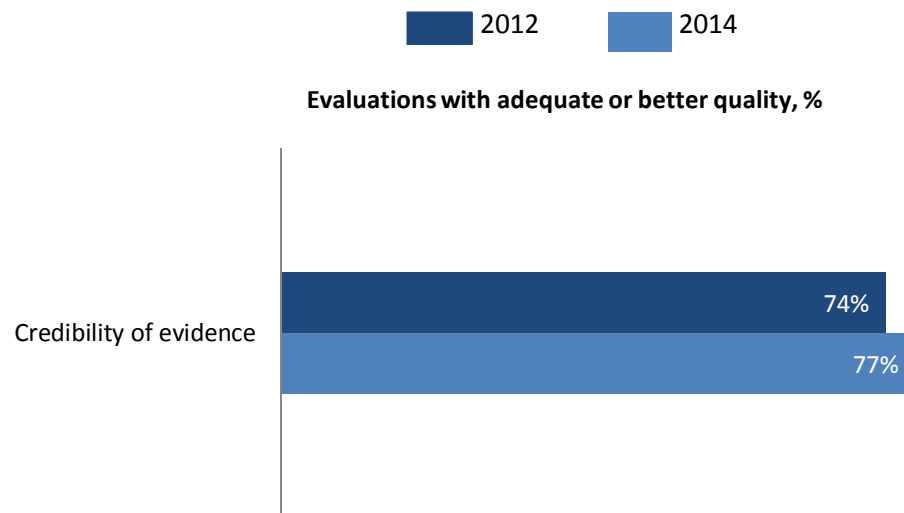


The quality of operational evaluations is good and they are a credible source of data for the aid program (6 of 8)

In addition to the findings discussed in previous slides, we chose the criteria 'credibility of evidence and analysis' to represent the overall quality of each operational evaluation. We chose this criteria because:

- Correlation analysis demonstrated there was a strong positive relationship between this and the other eight criteria
- This criteria focuses on the logical flow and evidence base of an evaluation. Using a common sense test, this is a good marker for overall evaluation quality
- The 2012 Review used this criteria to represent overall quality of each operational evaluation. Using it again allows for easier comparison between the current and previous Reviews.

77% of operational evaluations in the current Review were assessed as adequate or better quality for credibility of evidence and analysis. 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 assessed as adequate or better quality.



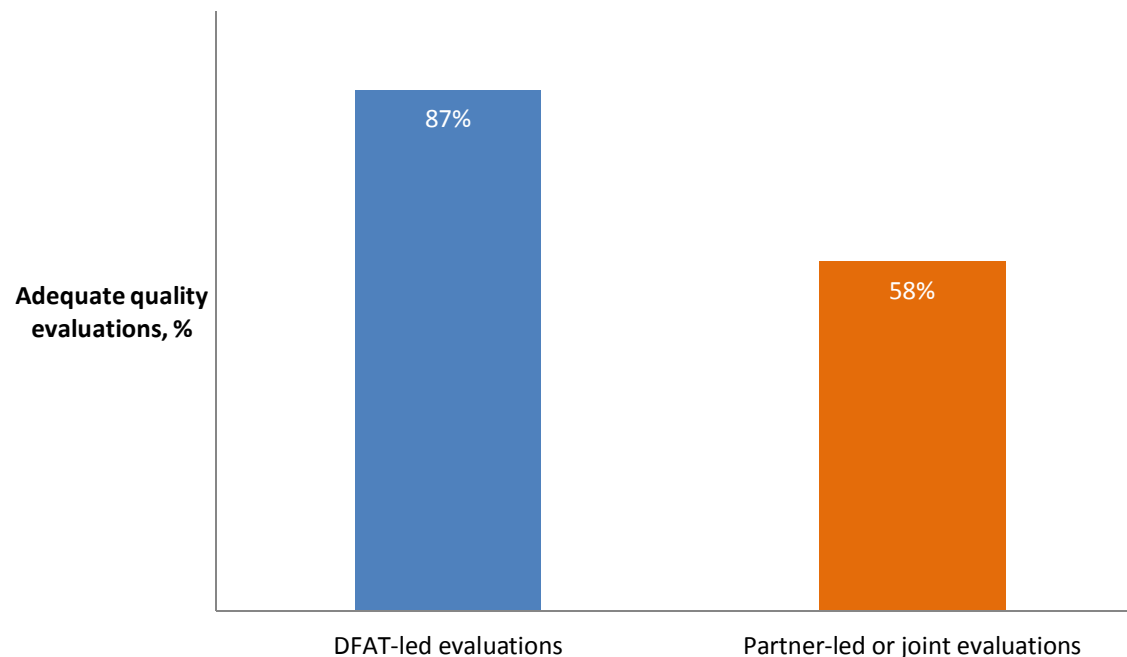
The quality of operational evaluations is good and they are a credible source of data for the aid program (7 of 8)

We compared the quality of evaluations that were managed solely by DFAT with evaluations which were led by one of DFAT's development partners, or were jointly managed by DFAT and a partner.

- A higher percentage of DFAT-managed evaluations were adequate or better quality, as shown in the chart below.
- This may be because, for partner-led/joint evaluations, DFAT devolves responsibility for evaluation quality to our partners and does not engage as actively when compared to DFAT-led evaluations.

As discussed in earlier slides, partner-led/joint evaluations offer opportunities to reduce the costs of operational evaluations.

- However, the lower quality of partner-led/joint evaluations raises questions about whether these approaches deliver value for DFAT.



Summary: The quality of operational evaluations is good and they are a credible source of data for the aid program (8 of 8)

In summary:

- Our assessments of operational evaluations against the nine quality criteria
- The comparison of evaluation quality in 2014 and 2012, and
- Our assessment against the proxy for overall evaluation quality, the criteria ‘credibility of evidence and analysis’

lead us to conclude that the quality of operational evaluation is good and they are a credible source of data for the aid program.

However, the quality of partner-led or joint evaluations, when compared to the quality of DFAT-managed evaluations, raises questions about the value delivered by partner-led/joint evaluations. DFAT should engage more actively with these evaluations to ensure our quality expectations and needs are met.

Good practice example

The evaluation plan and evaluation report for the “Review of Australian aid initiatives in the Pacific aimed at ending violence against women” are high quality evaluation documents which provide credible data.

The evaluation plan outlines a strong methodology. Data collection and analysis methods are well described and the limitations to the methods are clear and credible. Ethical issues are well handled, which is very particularly important given the challenges of working with survivors of violence.

The evaluation report demonstrates credible evidence and analysis. Data from international literature and informant interviews are interwoven to provide compelling evidence, quotes are well used to emphasise key points and the evaluators’ judgements are clear throughout the report.

Contents

Background and methods

Key findings

- Operational evaluations use modest resources
- The quality of operational evaluations is good
- For evaluation quality, the key is ‘everything in moderation’
- There are questions about how well operational evaluations are used
- DFAT does not have sufficient evaluation capacity to allow large numbers of evaluations to be conducted in-house

Recommendations

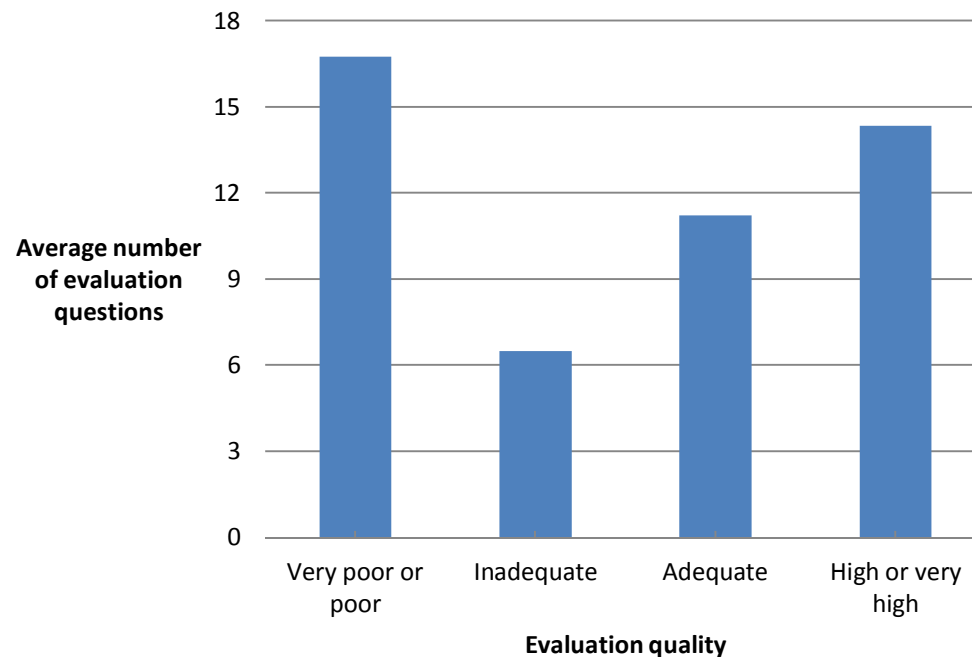
Evaluation snapshots and annexes

The third key finding is that for evaluation quality, the key is ‘everything in moderation’ (1 of 7)

To determine what contributes to evaluation quality, we examined several factors to see if there was any relationship between these factors and evaluation quality (measured using the criteria ‘credibility of evidence and analysis – see slide 25).

As shown in the chart below, operational evaluations rated as adequate or high quality had, on average, a moderate number of evaluation questions.

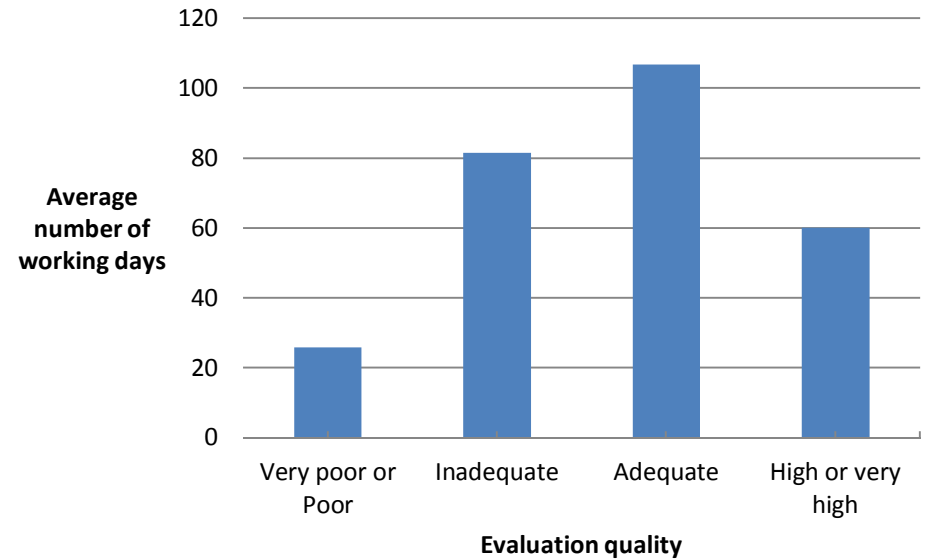
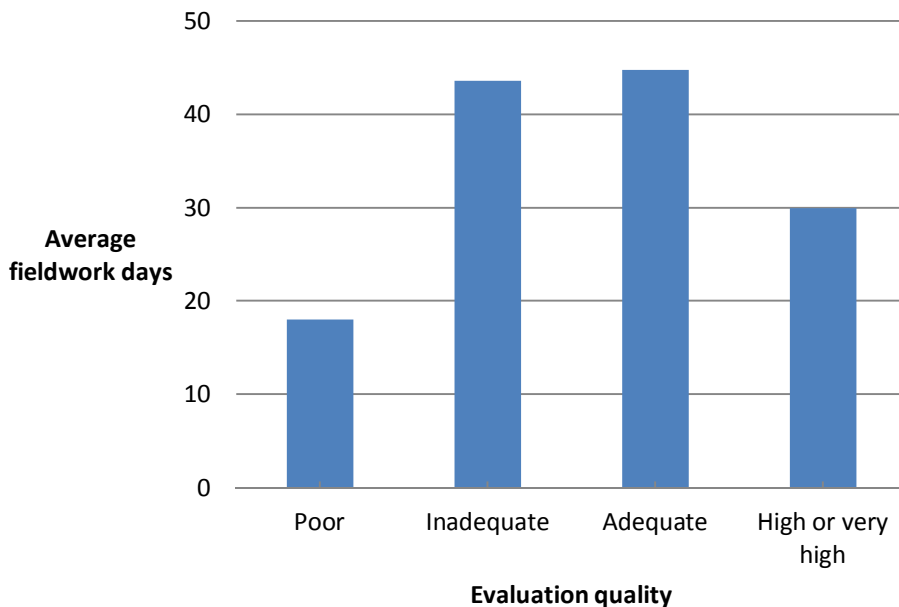
- Evaluations rated as inadequate or poor quality tended to have either a very large or very small number of evaluation questions.



For evaluation quality, the key is ‘everything in moderation’ (2 of 7)

Operational evaluations that had more working days were not necessarily better quality.

The evaluations rated as high or very high quality had, on average, fewer working days than evaluations rated as adequate or lower quality (see chart to the right).



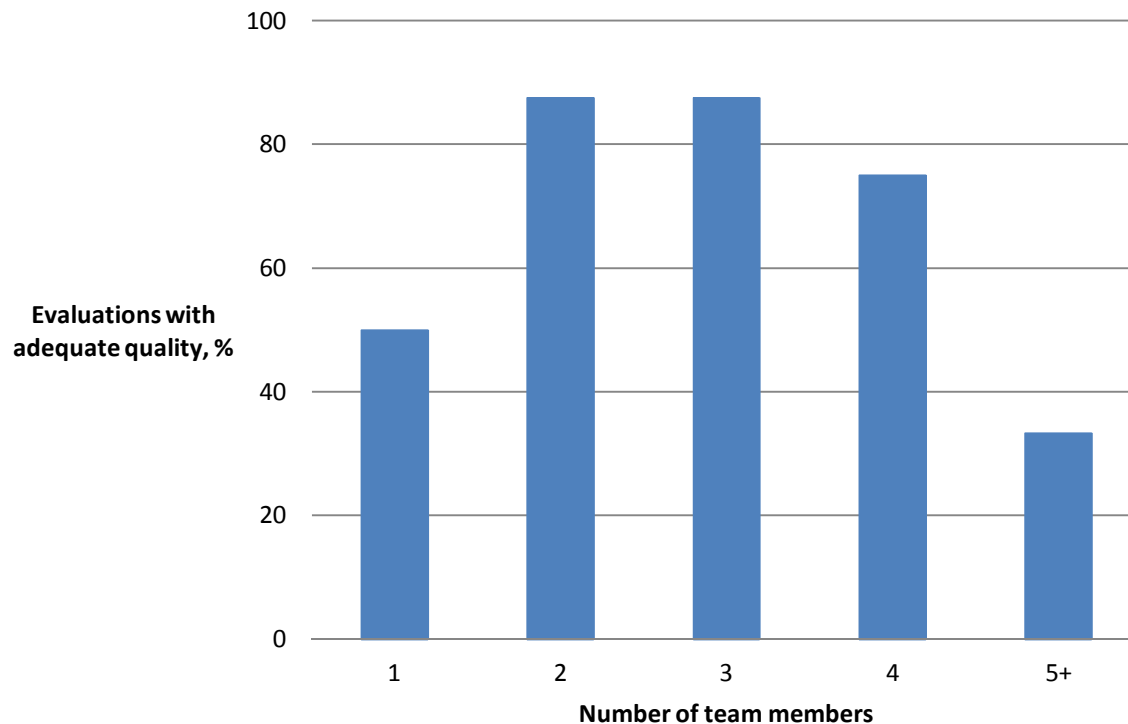
The same pattern holds for an evaluation’s fieldwork days: more fieldwork days did not necessarily lead to better evaluation quality.

The evaluations rated as high or very high quality had, on average, fewer fieldwork days than evaluations rated as adequate or lower quality (see chart to the left).

For evaluation quality, the key is ‘everything in moderation’ (3 of 7)

Larger team sizes were more likely to produce adequate quality evaluations, but only to a certain point:

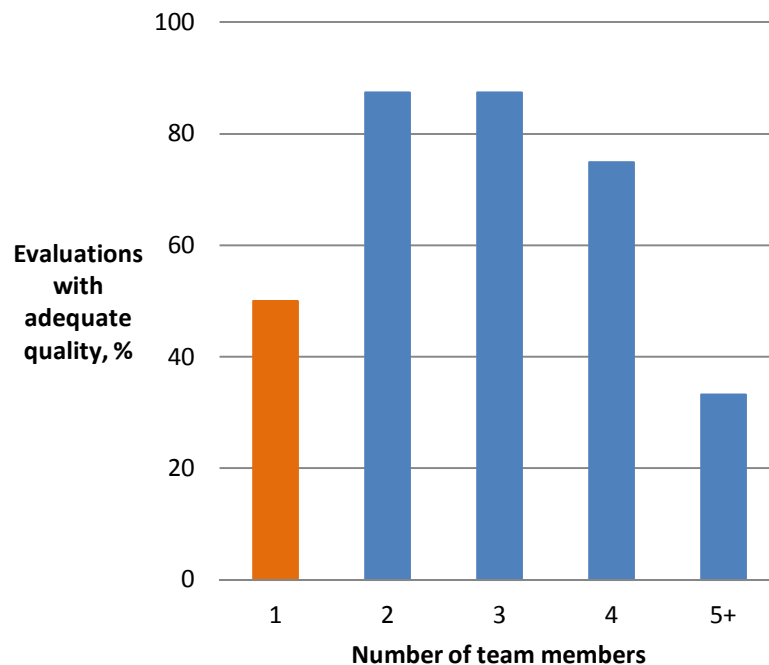
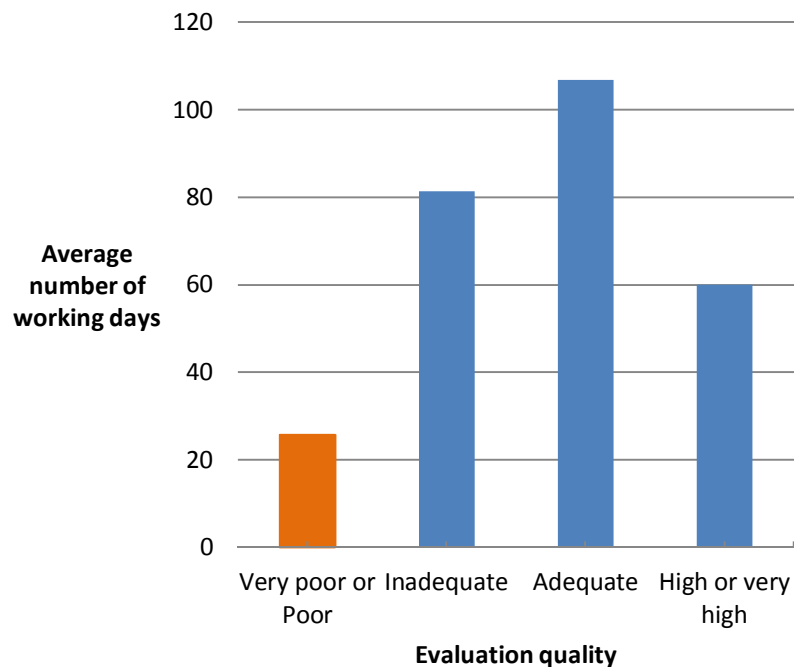
- As shown in the chart below, almost 90% of evaluations produced by teams with two or three members were adequate or better quality
- However, only a third of evaluations produced by teams of five or more were adequate or better quality
- This is consistent with the previous Review, which also found that very large evaluation teams produced fewer adequate quality evaluations.



For evaluation quality, the key is ‘everything in moderation’ (4 of 7)

In addition to the finding of ‘everything in moderation’, it is also clear that poorly resourced evaluations are much more likely to be low quality:

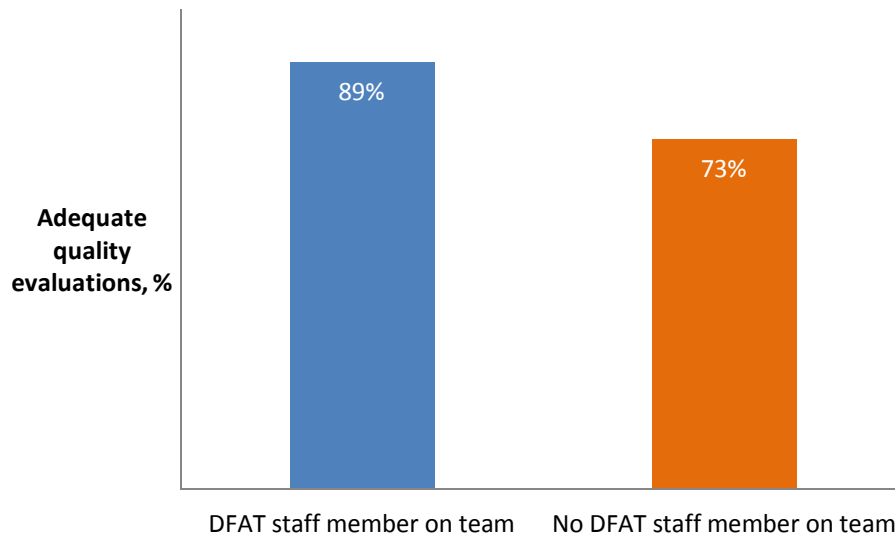
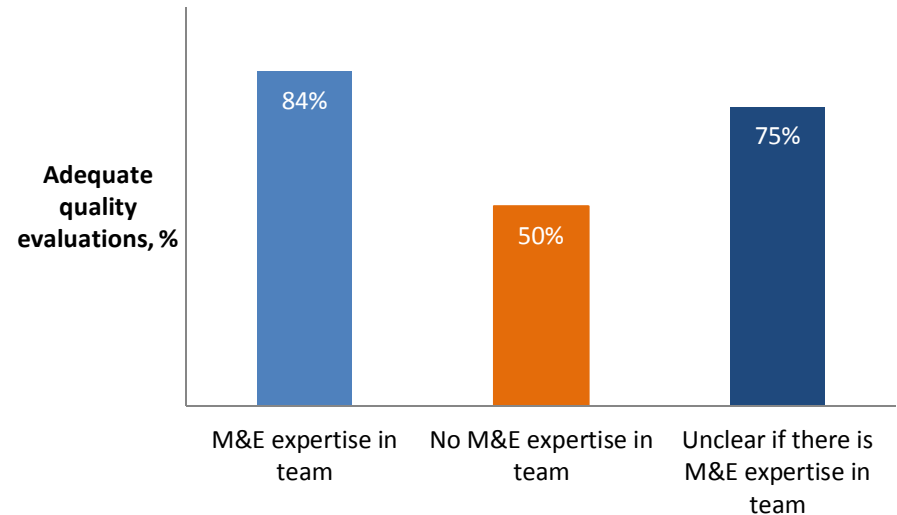
- As emphasised in the charts below (which are reproduced from previous slides), evaluations with a very low number of working days, or a very small team, were more likely to be poor quality
- This demonstrates that evaluations must be sufficiently resourced to increase the likelihood they will be adequate quality.



For evaluation quality, the key is ‘everything in moderation’ (5 of 7)

We also found that evaluation teams which included an M&E expert produced a higher percentage of adequate or better quality evaluations.

This is in comparison to evaluation teams which did not include an M&E expert, or when it was unclear if a team included an M&E expert (see chart to the right).

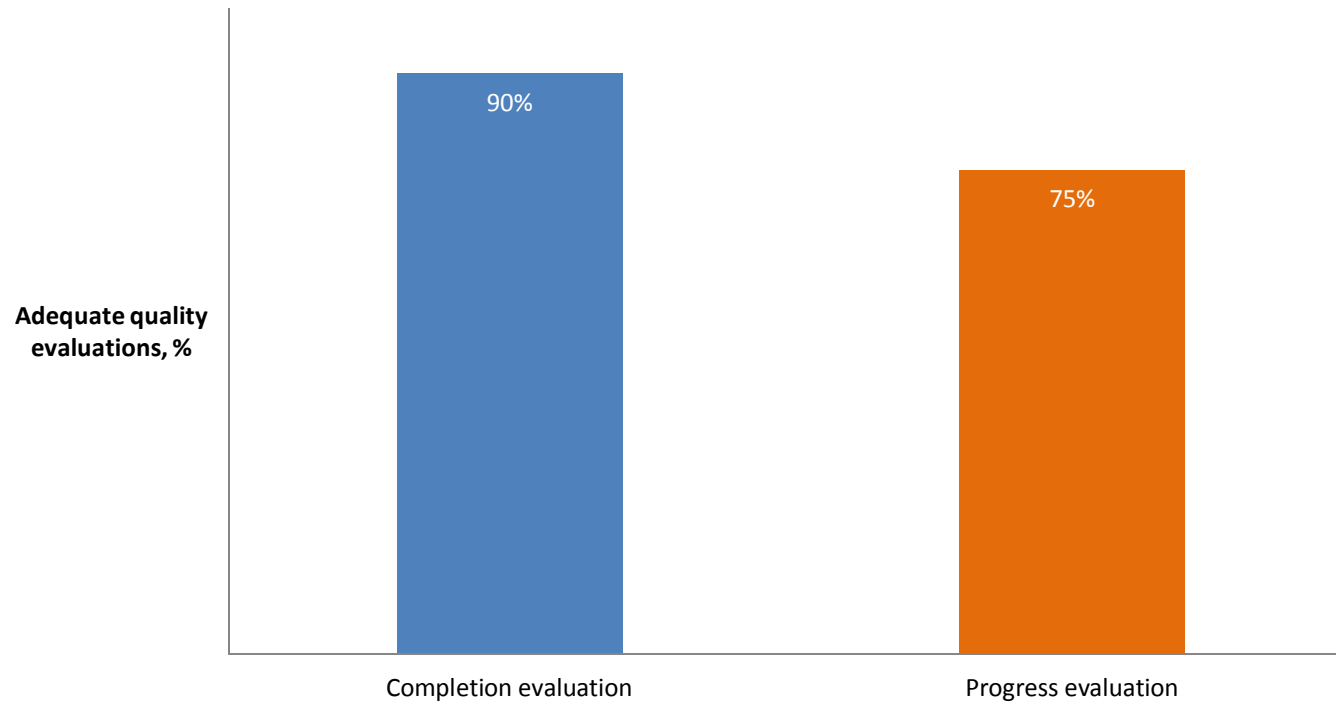


Evaluation teams which included a DFAT staff member also produced a higher percentage of adequate or better quality evaluations, when compared to evaluation teams which did not include a DFAT staff member (see chart to the left).

Note: Our sample only included nine evaluations with a DFAT staff member on the evaluation team, so this finding should be treated with some caution.

For evaluation quality, the key is ‘everything in moderation’ (6 of 7)

Finally, we found that a higher percentage of completion evaluations were adequate or better quality, when compared to progress evaluations, as shown in the chart below.



Note: Our sample only included 10 completion evaluations so this finding should be treated with some caution.

Summary: For evaluation quality, the key is ‘everything in moderation’ (7 of 7)

In summary, operational evaluations are more likely to be adequate or better quality if:

- They have a moderate duration, number of evaluation questions and team size, and
- The evaluation team includes an M&E expert.

Evaluations which include DFAT staff members may also be higher quality, while progress evaluations may require greater attention than completion evaluations to ensure they are good quality.

Evaluations with very low resourcing (for example, few working days or small teams) are much more likely to be low quality. This demonstrates that evaluations must have sufficient resourcing to increase the likelihood they will be adequate quality.

Good practice example

The terms of reference for the independent progress review of DFAT law and justice assistance in Indonesia provide a good practice example for evaluation planning.

The terms of reference outline a moderate scope for the evaluation with a reasonable number of prioritised evaluation questions. The resources dedicated to the evaluation, including team composition and size, number of working days and number of fieldwork days, appropriately match the scope. The role of each team member and the outputs the team must provide are also very clear.

Contents

Background and methods

Key findings

- Operational evaluations use modest resources
- The quality of operational evaluations is good
- For evaluation quality, the key is ‘everything in moderation’
- There are questions about how well operational evaluations are used
- DFAT does not have sufficient evaluation capacity to allow large numbers of evaluations to be conducted in-house

Recommendations

Evaluation snapshots and annexes

The fourth key finding is that there are questions about how well operational evaluations are used (1 of 5)

Although the quality of operational evaluations is good, the Review found there are questions about how well evaluation findings are used in DFAT.

- This concern about the use of evaluations is shared by many development organisations. The OECD Review of Evaluation Systems in Development Cooperation found that insufficient use of evaluation findings is an issue in the majority of the 49 organisations examined.

Evaluation use covers whether direct action is taken as a result of an evaluation, and whether lessons from an evaluation are made available and applied to other investments.

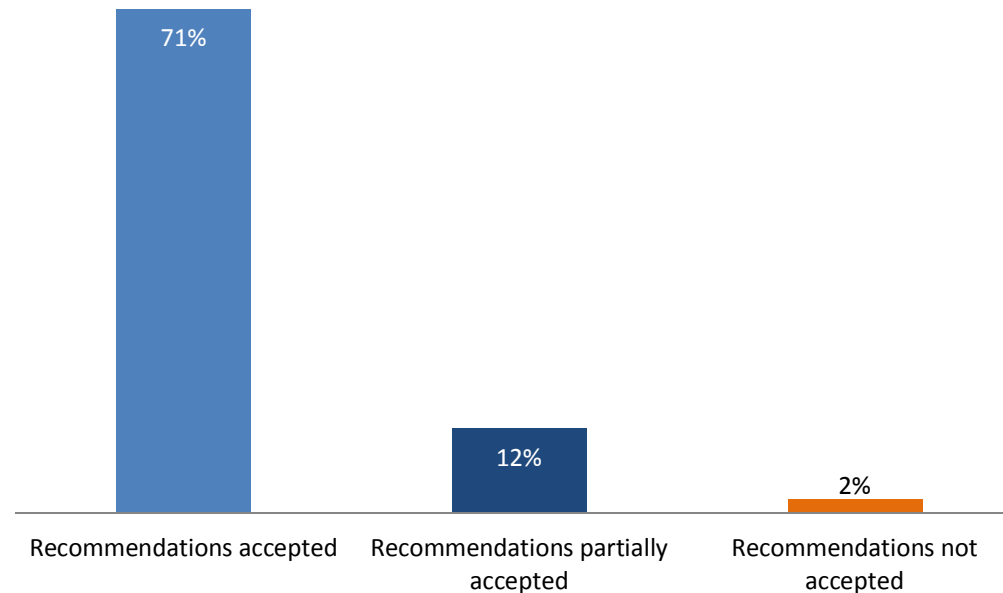
There are indications that direct action may not be taken as a result of all DFAT operational evaluations:

- Many evaluations did not appear to have management responses; we were only able to locate 16 out of a possible 35 management responses
- This is consistent with the experience of other donor organisations, with the OECD Review finding management responses are often not completed in a timely manner
- As noted in slide 24, only 66% of operational evaluations were assessed as having adequate or better quality recommendations. This may be influencing the low number of management responses.

There are questions about how well operational evaluations are used (2 of 5)

On a positive note, analysis of the 16 available management responses showed 71% of recommendations were accepted and 12% were partially accepted (see chart below).

- This implies at least some operational evaluations are being used to inform action on investments.

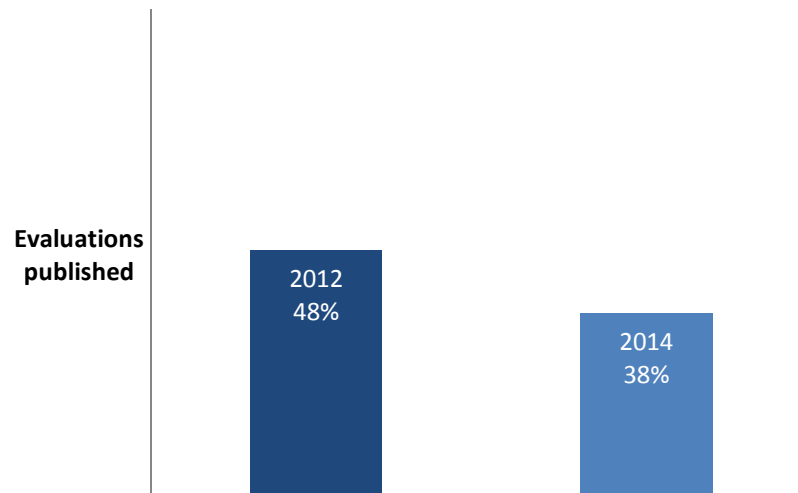


Note: DFAT's response to 15% of recommendations was unclear. This was for reasons such as the management response was general and didn't discuss specific recommendations. Such unclear responses are not represented in the chart above.

There are questions about how well operational evaluations are used (3 of 5)

Lessons from operational evaluations are not always made widely available:

- Although it is mandatory to publish completed operational evaluations, by July 2015 only 38% of operational evaluations completed in 2014 had been published
- This is consistent with the low publication rate identified in the previous Review, as shown in the chart below.



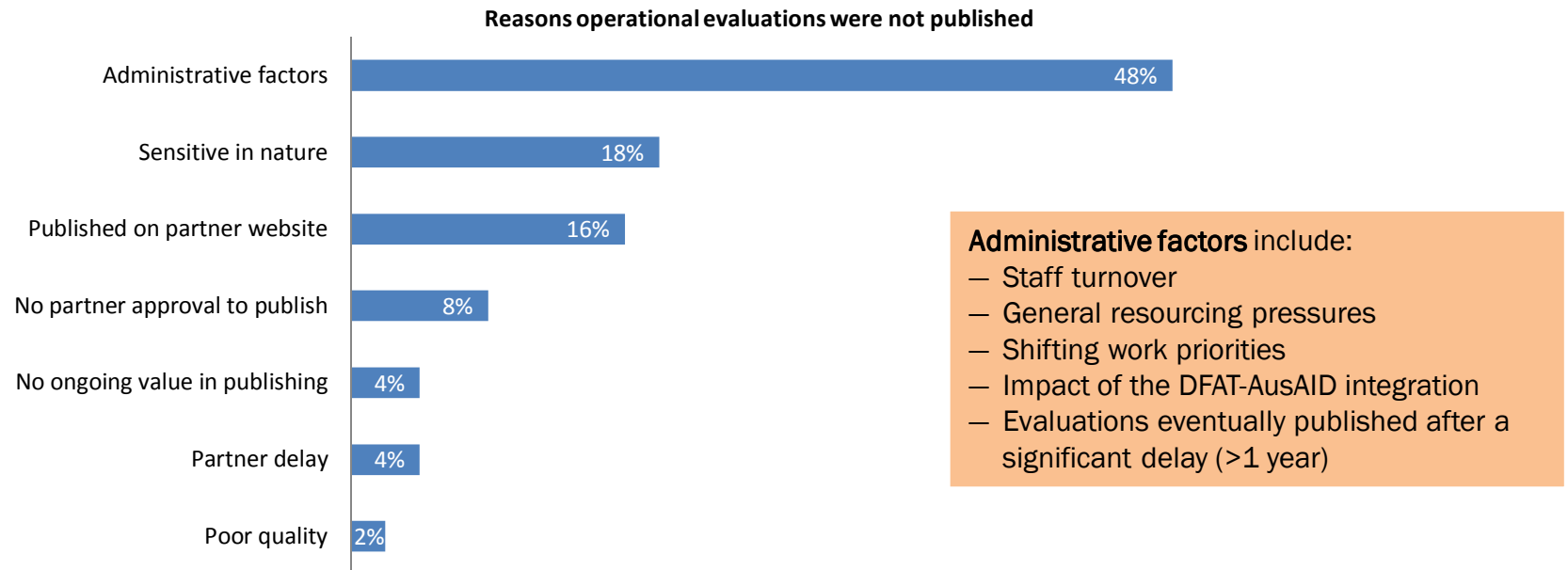
Notes: The publication rate of 38% is for all 77 operational evaluations completed in 2014. An additional 2 evaluations (3% of total evaluations) had been granted an exemption from publication by July 2015.

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.

There are questions about how well operational evaluations are used (4 of 5)

The DFAT Internal Audit Branch's audit of evaluation publication (2016) examined 50 operational evaluations which had been completed but not published on the DFAT website:

- It found that almost half of the evaluations were not published due to administrative factors, particularly staff resourcing constraints and high staff turnover
- Other reasons for non-publication include that the evaluation was too sensitive; the evaluation was published on a partner website but not on the DFAT website; and that a partner did not agree to publication (see chart below).



Summary: There are questions about how well operational evaluations are used (5 of 5)

In summary:

- The difficulty locating management responses
- The low publication rate, and
- The consistency between the findings of this Review and the experiences of other development organisations, as discussed in the OECD Review of Evaluation Systems in Development Cooperation

lead us to conclude there are questions about how well operational evaluations are used in DFAT.

Good practice examples

The management response to the review of the Enterprise Challenge Fund Pilot Program provides a good practice example. The management response includes a good summary of the evaluation as well as a credible critique of the evaluation's weaknesses. It also demonstrates how recommendations can be handled when an investment has been closed.

A second good practice example is the management response to the mid-term review of the Australia-Africa Community Engagement Scheme. The management response is strong because the actions to address each recommendation, and those responsible for doing so, are specific and clear. It is also a good example of how a management response can be completed when an investment has multiple implementing partners.

Contents

Background and methods

Key findings

- Operational evaluations use modest resources
- The quality of operational evaluations is good
- For evaluation quality, the key is ‘everything in moderation’
- There are questions about how well operational evaluations are used
- DFAT does not have sufficient evaluation capacity to allow large numbers of evaluations to be conducted in-house

Recommendations

Evaluation snapshots and annexes

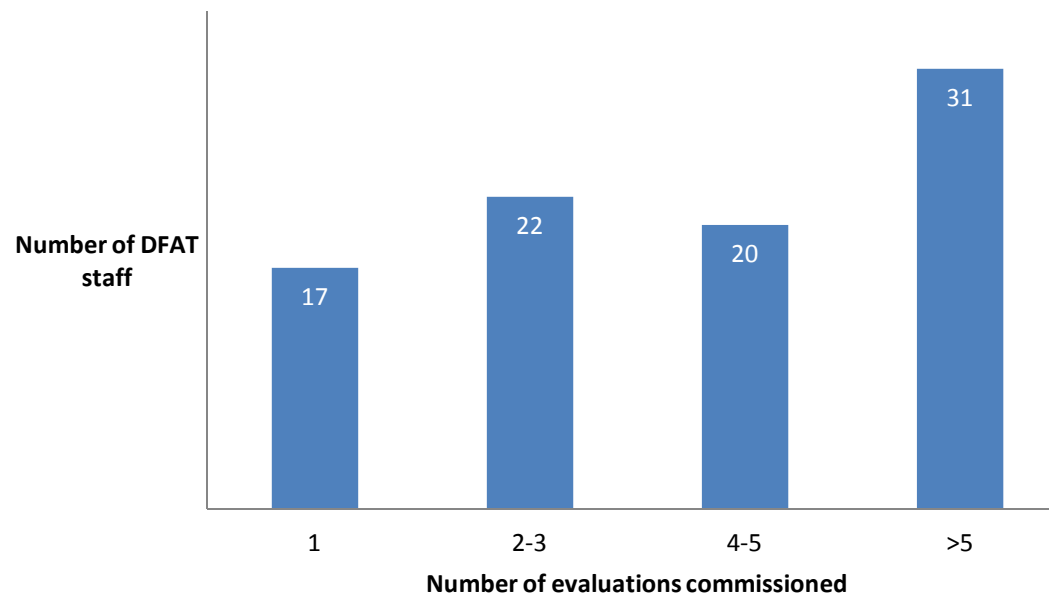
The fifth key finding is that DFAT does not have sufficient evaluation capacity to allow large numbers of evaluations to be conducted in-house (1 of 4)

As outlined in slide 11, we conducted a short all-staff survey to assess DFAT's capability to commission and conduct operational evaluations.

- 113 staff responded to the survey.

The survey results show that DFAT has some evaluation capability:

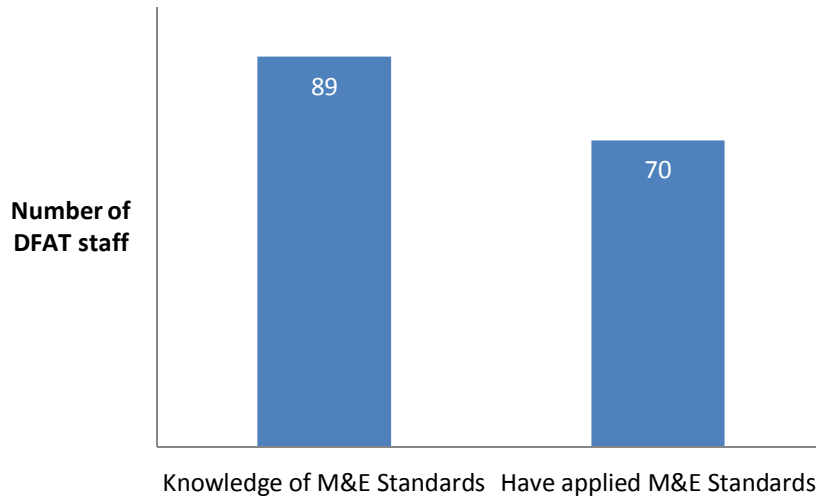
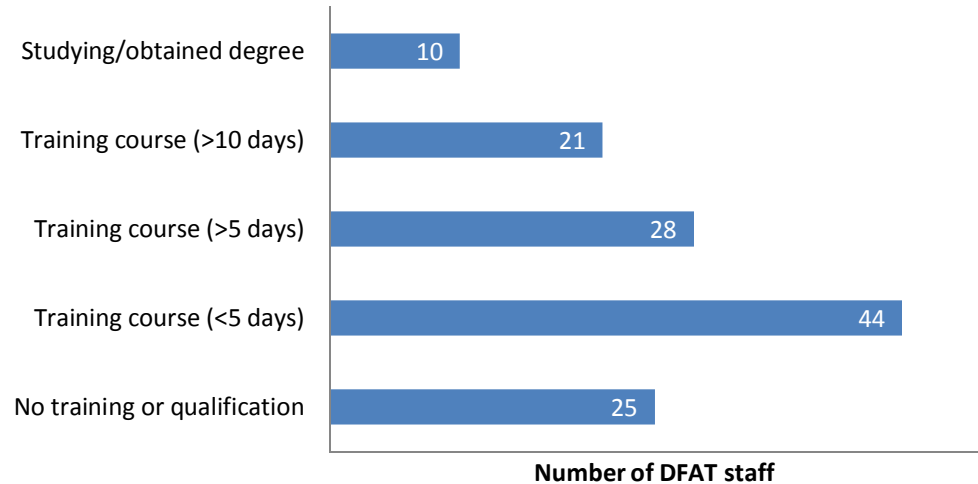
- As shown in the chart below, 90 respondents (80%) have commissioned an evaluation
- The largest group (31 respondents) have commissioned more than five evaluations.



DFAT does not have sufficient evaluation capacity to allow large numbers of evaluations to be conducted in-house (2 of 4)

88 respondents (78%) have had some training in evaluation (see chart to the right).

The largest group (44 respondents) has had up to 5 days training.



89 respondents (79%) have knowledge of DFAT's Monitoring and Evaluation Standards (see chart to the left).

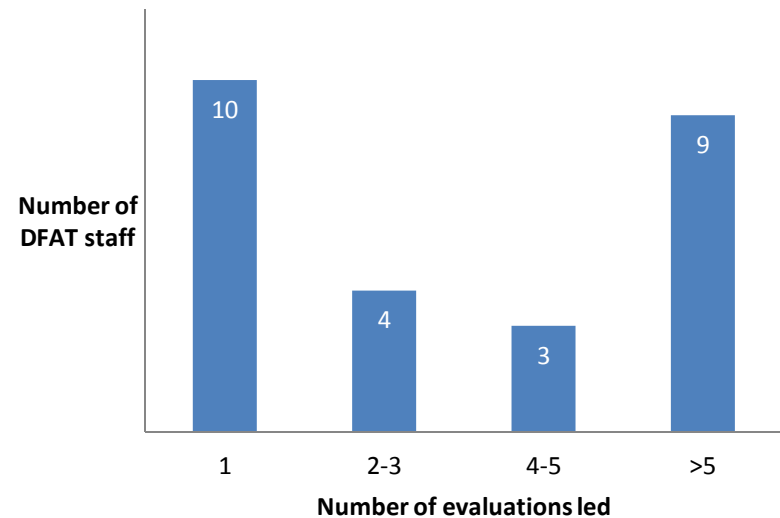
70 respondents (62%) have experience applying them.

DFAT does not have sufficient evaluation capacity to allow large numbers of evaluations to be conducted in-house (3 of 4)

However, only 57 respondents (50%) have played a substantive role in an evaluation team, such as contributing to an evaluation plan, collecting data, or writing parts of an evaluation report.

In addition, only 26 respondents (23%) have led an evaluation.

- Of this group, the largest number have led only one evaluation, as shown in the chart to the right
- Our analysis shows staff who have led an evaluation are widely dispersed across the department, working in 14 of DFAT's 24 divisions.



DFAT's issues with evaluation capacity are consistent with other development organisations:

- The OECD Review of Evaluation Systems in Development Cooperation found many organisations have gaps in their evaluation capacity, both in central evaluation units (such as ODE) and in staff who manage operational evaluations
- The OECD Review also found that two thirds of the organisations examined outsourced at least 75% of their evaluations to external consultants.

Summary: DFAT does not have sufficient evaluation capacity to allow large numbers of evaluations to be conducted in-house (4 of 4)

In summary:

- DFAT has some evaluation capability. Staff have accessed training, have experience commissioning evaluations and also have knowledge and experience applying the DFAT M&E Standards
- However, few staff have played a substantive role in an evaluation team or have experience leading evaluations.

This leads us to conclude that DFAT does not have sufficient evaluation capacity to conduct a significant number of evaluations in-house without the support of external consultants.

Contents

Background and methods

Key findings

- Operational evaluations use modest resources
- The quality of operational evaluations is good
- For evaluation quality, the key is ‘everything in moderation’
- There are questions about how well operational evaluations are used
- DFAT does not have sufficient evaluation capacity to allow large numbers of

Recommendations

Evaluation snapshots and annexes

The Review provides recommendations for IEC consideration

As discussed in the limitations section (slide 12), the Review provides recommendations for the IEC's consideration.

- As ODE oversees DFAT's operational evaluations policy, ODE would need to provide a management response to any recommendations which are made.
- Given ODE is also writing this Review, it is impractical for ODE to draft recommendations which we then have to respond to.
- However, it is useful for the IEC to review recommendations and note whether they flow logically from the data and should be supported.
- If the IEC supports the following recommendations, ODE will then act on them.

DFAT has decided to further simplify the operational evaluations policy

As ODE was finalising this Review, DFAT was considering measures to further simplify management of the aid program.

- In November 2015, DFAT decided that the operational evaluations policy should be revised again, with a view to ensuring the time and resources required to produce evaluations is commensurate with the benefits DFAT receives from them.
- ODE has been tasked to develop the revised policy.

The findings of this Review highlight three recommendations that need to be addressed in the update of the operational evaluations policy, as discussed in the following slides.

Recommendations: Improving the use of operational evaluations

The updated operational evaluations policy needs to encourage greater use of operational evaluations.

- It is clear from the Review’s findings that operational evaluations are not being used to their full potential
- They can be better used to inform direct action on investments, and to share knowledge more broadly across DFAT and the aid community.
- More robust mechanisms and oversight are also need to ensure management responses are completed and operational evaluations are published on the DFAT website.

To improve the use of operational evaluations, it is recommended that:

- 1) DFAT’s Executive has the opportunity to consider planned, completed and published evaluations. This should be provided through regular reporting from ODE.
- 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.

Recommendation: Improving DFAT's evaluation capacity

It is clear from the Review's findings that DFAT's evaluation capacity is insufficient to allow large numbers of evaluations to be conducted in-house.

- At the same time, the Review also tentatively found that evaluation quality may be better when DFAT staff are included on evaluation teams.
- Experience also tells us that evaluation findings are more likely to be used when DFAT staff have strong ownership of them.

DFAT will need to continue to use consultants to conduct evaluations in the short term. However, in the long term DFAT should aim to improve the evaluation capacity of its staff. This will not only reduce reliance on consultants; it is also likely to improve the quality of evaluations, particularly the quality of recommendations, and the use of evaluation findings by DFAT.

To address these points it is recommended that:

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

Contents

Background and methods

Key findings

- Operational evaluations use modest resources
- The quality of operational evaluations is good
- For evaluation quality, the key is ‘everything in moderation’
- There are questions about how well operational evaluations are used
- DFAT does not have sufficient evaluation capacity to allow large numbers of

Recommendations

Evaluation snapshots and annexes

Evaluation snapshots

To facilitate learning and knowledge sharing, we identified useful lessons on aid and development from the operational evaluations which we examined.

- These lessons are summarised as evaluation snapshots in Part 2 of this report.

The lessons are:

- Gender equality and working with the poorest: Lessons from the Laos Australia NGO Cooperation Agreement Program
- Supporting the private sector: Lessons from the Enterprise Challenge Fund for the Pacific and South East Asia
- The challenges of cascade training: Lessons from the Malaysia Australia Education Project for Afghanistan
- Assessing value for money: Different methods from three operational evaluations
- The perils of minimising costs: Lessons from the Regional HIV Capacity Building Program.

Annexes

The following annexes are available in Part 3 of this report:

- Annex A: Review methods
- Annex B: List of operational evaluations included in the Review

References

Danish Management Group, 2016. *OECD Review of Evaluation Systems in Development Cooperation*. Draft report, February 2016.

Department for International Development (DFID), 2014. *Rapid Review of Embedding Evaluation in UK Department for International Development*. Available at <https://www.gov.uk/government/publications/review-of-embedding-evaluation-in-the-department-for-international-development>

Department of Foreign Affairs and Trade (DFAT), 2016. *Internal Audit 2015-16: Publishing of Independent Aid Project Evaluations* (internal document).