

DFAT Review of operational evaluations 2014



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.

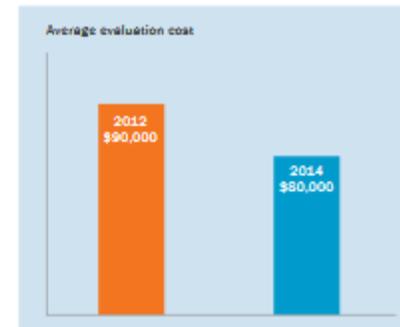


Chart 1: The average cost of operational evaluations decreased between 2012 and 2014

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.

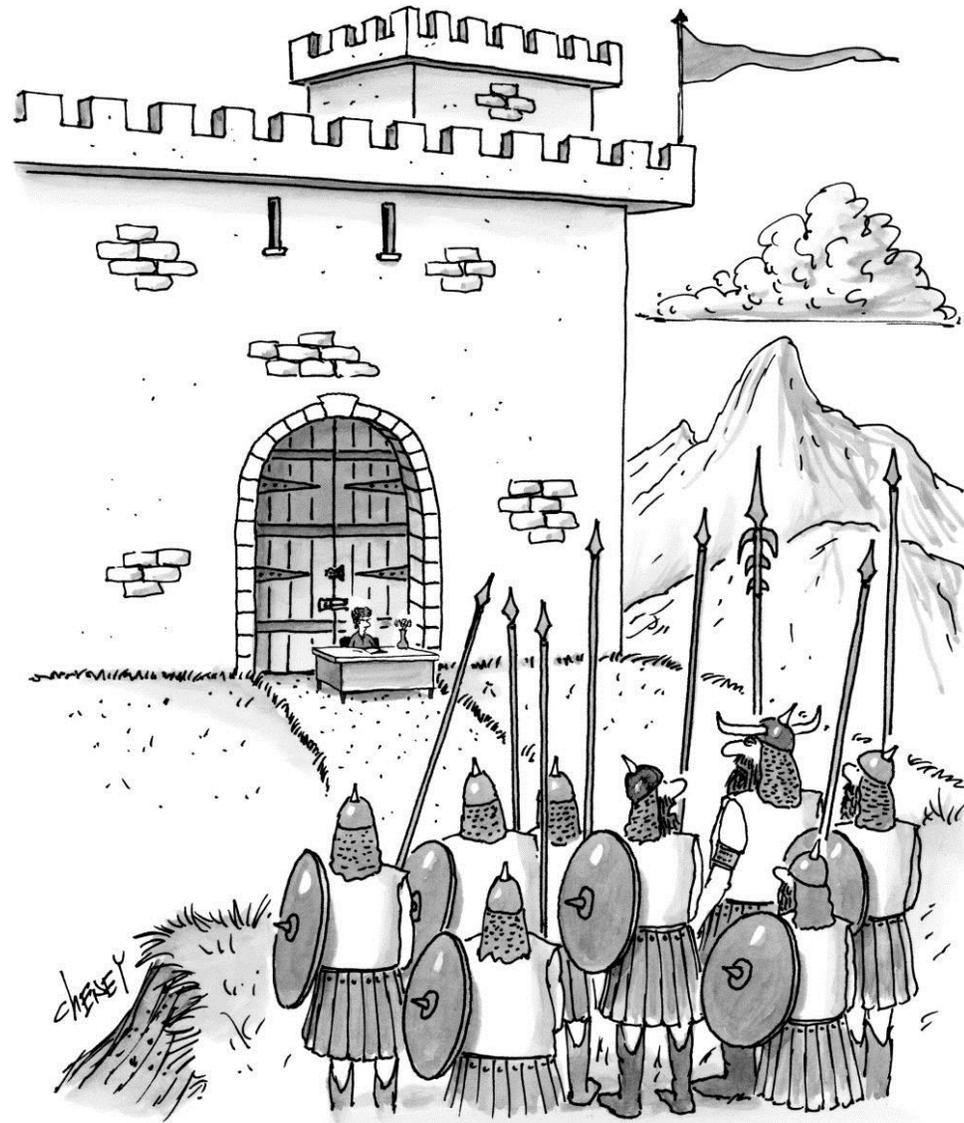
The proportion of operational evaluations conducted jointly by DFAT and one of our partners, or led by one of our partners, also increased from 17% in 2012 to 34% in 2014. In addition, the proportion of cluster evaluations, or evaluations which examine more than one investment, increased from 7% in 2012 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.

Discussant: Terence Wood
(terence.wood@anu.edu.au)

The plan:

1. My interpretation of findings
2. My findings about how they found their findings

What follows is definitely not an attack!



Key findings (as I see them):

Most DFAT evaluations are quite good

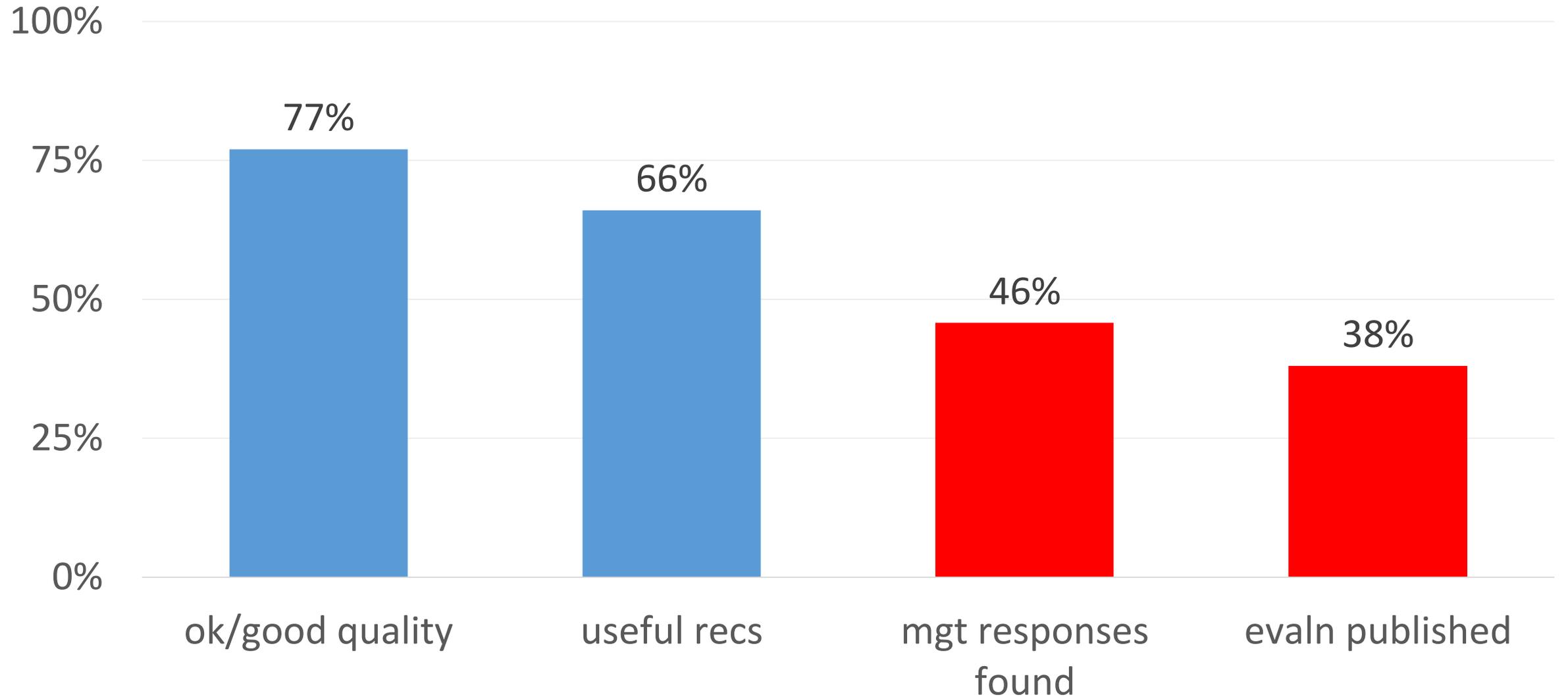
Evaluations being thought about intelligently in DFAT (e.g. this report)

But DFAT does not invest nearly enough time nor people-power into the evaluation process

- average evaluation cost as a proportion of activity value 0.37% - are we really *that* confident that aid works? (DFID 1.9%)
- evaluation attrition

Evaluation attrition

2014 - % of evaluations surveyed (blue); % of evaluations (red)



Methodology! Findings about how they found their findings



"I've been working out for 6 months but all my gains have been in cognitive function."

This study is good!:

- convincing enough and important points identified
- methods are spelled out (and limitations noted; albeit in an appendix)
- coding/scoring schema made sense

But in future...

- using the population or increase the sample (two years?)
- **Careful with causal claims**
- do evaluations have an impact?
- **in-depth case studies (particularly of failure)**

An example where causal care was needed

No reason to believe “high or very high” quality evaluations are a “high or very high” because they have fewer working days. It could be that many ‘adequate’ evaluations would have been worse with fewer days. Maybe in some cases, more days were needed cause it was a more difficult job. “Maybe high or very high” evaluations were good and quick because you had really good evaluators.

