





		Reality	
		Program doesn't work	Program works
Research	Program doesn't work	Accurate 	Type I error 
	Program works	Type II error 	Accurate 

ARDE 2009: Independent aid evaluation?

By Matthew Morris

This post for the Aid Open Forum by Matt Morris, Deputy Director of the Development Policy Centre, concludes our series of articles on the 2009 Annual Review of Development Effectiveness ([ARDE](#)). The other articles in the series can be found [here](#).

Evaluation reports, such as the ARDE, are especially important for an aid program that has doubled in the last five years and is set to double again to \$8.6 billion by 2015-16.

I've enjoyed reading both the ARDE - whose findings can be found [here](#) - and the discussion by [Peter McCawley](#), [Ian Anderson](#) and [Richard Curtain](#).

First, and most of all, I'd like to commend the Office of Development Effectiveness

(ODE) and AusAID for producing the third ARDE report. A huge amount of work – involving all of AusAID’s programs – goes into the various self-assessments that build up to the final report, and this is part of the broader performance orientation of the aid program. As a recent [ANAO review](#) noted:

AusAID has: implemented a robust performance assessment framework for aid investments; commenced valuable annual program reporting; strengthened its quality reporting system for aid activities; and established ODE to monitor the quality and evaluate the impact of Australian aid. These efforts are focusing agency attention on the quality of country programs and aid activities, and the factors that lead to better development outcomes.

As noted in the three sets of comments, the 2009 ARDE provides some useful insights and analysis on the performance of the aid program (e.g. the fragmentation discussion) that make the report important reading for anyone interested in Australian aid.

Peter McCauley notes that the report is often ‘gently critical’, such as the ‘thrashing with a feather’ of traditional approaches to aid. Though in fairness not all of the tone is soft – the discussion of fragmentation is pretty hard – but certainly parts are and I agree with Peter’s characterisation for the most part.

Making confident choices

Like all good economists, I am frugal and cautious (though I’m not sure which way the causality runs on this.) These are good habits for thinking about scaling up aid, and economists have a range of tools and concepts to guide discussions.

One useful concept is Type I and Type II errors. This refers to two types of errors that can be made in hypothesis testing that can lead to incorrect conclusions. If a null hypothesis is incorrectly rejected when it is in fact true, this is called a Type I error (also known as a false positive). A Type II error (also known as a false negative) occurs when a null hypothesis is not rejected despite being false.

In the context of scaling up aid, we want to be sure that we don't scale up programs that don't work (a Type I error) or fail to scale up programs that do work (Type II error).

High quality evaluation - timely, rigorous and policy relevant - is an key tool for helping policy-makers avoid these kinds of errors.

The cost of delay

Yet it is because of the importance of evaluation, and all of the work that has gone into the ARDE and its high quality of analysis, that I am disappointed with a key failing - it's tardiness. The late and low-key publication of the report potentially diminishes the ARDE's impact on programs and development policy, and this raises several issues that deserve further discussion.

First, if the ARDE is intended to provide feedback into program choices, then ideally its publication should be synchronised with key decision points (e.g. in the annual budget timetable.) For example, a report on 2008-09 would need to be available by December 2009, around the time that budget submissions are being made, so that it's findings can be factored in to the 2010-11 budget.

Secondly, the late publication of the report means that recent program and policy developments are missed out. As Ian Anderson points out, an analysis of Copenhagen and it's aftermath is missing from the discussion on climate change. Another missing piece of the jigsaw is all the work AusAID has done recently to [review technical assistance](#). Technical assistance is highlighted in the report as an issue, but there is no mention of the work underway to move away from excessive reliance on advisers. Indeed, the late publication of the report may incorrectly create an impression no action is being taken.

Thirdly, these delays and gaps could be managed, to some extent, through the publication of an AusAID management response - setting out how the report's recommendations were taken forward, including through the 2010-11 budget.

This could have filled some of the gaps mentioned above, but more importantly it was a key recommendation of the ANAO's report on ['AusAID's Management of the Expanding Australian Aid Program'](#).

The evaluations need to be improved, however, by inclusion of management responses, as is required for independent evaluations of activities. Currently, the only articulated response to the findings of the ARDE is ODE's own forward agenda. It is not apparent how the agency as a whole intends to respond, nor is it clear how this response will be coordinated. Management responses to ODE evaluations would increase ODE's leverage with respect to program management and more clearly delineate its responsibilities from program delivery areas. (p.139)

Fourthly, the delay in the publication of the ARDE also raises questions about the appropriateness of the in-house evaluation model for AusAID. Two of the big benefits of the in-house model are supposed to be that it is easier for frank and informed discussions to take place and lessons to be passed through into programs and policy. Both of these may be happening, but the soft tone of the ARDE, the report's delay and the lack of a management response make it difficult to know.

The big drawback of the in-house model is that assessments can lack objective rigor and hard choices may be avoided. The low proportion (6%) of programs not meeting their objectives and the lack of a management response (e.g. on what the agency is doing about these programs) raise questions about both of these possible weaknesses.

Implications for the Aid Review

The Independent Review of Aid Effectiveness will be looking closely at the findings of the 2009 ARDE as part of their assessment of the last doubling of Australian aid, but they will also need to get much more information from ODE

and AusAID to bring the official assessment up to date.

As part of the next doubling of aid, the Review Panel should also consider how evaluation can help Australia to scale up aid programs that work and feedback lessons into program design and implementation.

It is these kind of issues that have led some other donors, notably the UK, Sweden and the World Bank, to move to a more independent model of evaluation. The UK government is establishing an [Independent Commission for Aid Impact](#) (ICAI) and Swedish Agency for Development Evaluation ([SADEV](#)) already exists the umbrella authority evaluating all of Sweden's development assistance work. And the World Bank's [Independent Evaluations Group](#) carries out a range of evaluations, including an [Annual Review of Development Effectiveness](#), reporting directly to the Bank's Executive Board. Are there lessons for Australia?

Perhaps we need a debate on independent evaluation of the Australian aid program.

Matt Morris is a Research Fellow at the Crawford School and Deputy Director of the [Development Policy Centre](#).