

## The need for more rigor in AusAID's project evaluations

**Author :** Dinuk Jayasuriya

**Date :** May 15, 2012



Aid effectiveness has become the mantra of the day. Aid donors demand it, program implementers profess to embody it and consultants and companies make money assessing it. AusAID is all about it, even announcing an Independent Evaluation Committee and a Results Framework as part of the aid budget.

Within an organisation tackling aid effectiveness takes many forms. There's the monitoring which at a basic level involves carefully setting SMART (Specific, Measurable, Achievable, Relevant and Time-bound) objectives and then gathering data periodically to meet targets. Then there are evaluations which are a completely different beast altogether.

Evaluations can be designed by an independent body; the Independent Evaluation Groups of the [World Bank](#) and [Asian Development Bank](#) are such examples. For AusAID, this function is performed by the [Office of Development Effectiveness](#) (ODE).

Similarly to the Multilateral Banks, ODE undertakes [evaluations](#) of country programs and sectoral evaluations in areas such as public sector reform, health, education and water and sanitation. However, unlike the multilaterals, there appears to be little in the way of *independent* evaluations at a project level. The recently released AusAID 'Blue Book' relating to the budget shows this may change with ODE tasked with undertaking program level evaluation. Any omission of such evaluations is worrying if Jeffrey Sachs is to be [believed](#); "if someone is really serious about the aid issue they should take a micro-perspective" to evaluation.

Evaluations can also be undertaken by project teams themselves (as distinct from ODE) with support from internal or external evaluation experts. Unfortunately, AusAID project evaluations, are (to the best of my knowledge) *all* mid-term reviews or *ex-post* (i.e. undertaken after the program has completed) which could be one symptom of evaluations being an after-thought rather than an institutionalised and integral part of the learning process. In any case, what's missing is evidence that an evaluation has been designed *ex-ante* (i.e. before the project was implemented). While country and sectoral evaluations are difficult to design in advance, this is not the case for project evaluations.

Why is this important? Consider a project intervention that involves providing text-books to students in a poor region of Pakistan. Assume we calculate the test scores of students before the project starts and notice after it finished that student scores have improved. Is that a big tick for our aid program? Only if we can successfully attribute the increase in scores to the text-books. What if the increase was due to these students receiving extra tutoring or having access to a sudden influx of higher qualified teachers? Also what if these students were unusually intelligent. If they weren't, would extra text-books have made a difference?

By considering an evaluation *ex-ante*, a technical expert can incorporate what us evaluators like to call counterfactuals. Moreover, we can randomly assign a project intervention (eg: text-books) to some schools

in the district and not to others. Assuming a large enough sample size, the average school that receives the project should have the same characteristics as the average school that doesn't receive the project. Any difference after project completion can, on average, be attributed to the intervention (i.e. the text-books).

The above example is a very simplistic example of a randomized evaluation. In many cases such evaluations are [inappropriate](#) or impractical. They require large sample sizes, regular monitoring and specific interventions (for example, they may not be suitable for governance programs). They have been criticised for not being universally applicable and for focusing too much on quantitative analysis. Assessing long-term impacts could also be problematic. Most randomized evaluations are undertaken just after project completion. Thus they are likely to estimate average quantitative *outcomes* rather than impacts.

I agree with these criticisms of randomized evaluations and propose that where practical, the most rigorous evaluations of outcomes have to follow a mixed methods prescription that incorporates a randomized component and a battery of qualitative tools. Let's call such evaluations the 'platinum standard'. A randomized evaluation (the gold standard) by itself is unlikely to be enough; just as a mixed methods approach without a randomized trial is unlikely to be enough.

Following an ex-post evaluation approach with largely qualitative analysis (similar to the approach that AusAID takes for most of its projects) can [overstate](#) aid effectiveness. To be fair, very few organisations undertake randomized evaluations on an institutional and large scale basis. However they can be done. The [MIT Poverty Action Lab](#) has undertaken more than 300 randomized evaluations, the World Bank has a separate [unit](#) dedicated to just doing randomized evaluations and DFID and USAID also commission some randomized evaluations. These organisations can go one step further by following a platinum standard prescription combining randomized evaluations with qualitative analysis.

I've been informed AusAID is working with [3IE](#) to develop a randomized control trial of a program/project in Indonesia. This is commendable and important; but a token evaluation is not enough. If AusAID wants to be the best donor possible, it needs to administer platinum standard evaluations in at least *some* (not one!) of its projects. The following are possible (albeit overly simplistic) steps forward for evaluators involved in country programs (as distinct from ODE):

1. Have an evaluation expert investigate AusAID's pipeline projects and select all projects that are appropriate for platinum standard evaluation.
2. Randomly select projects for such an evaluation.
3. Select an external partner to implement the platinum standard evaluation.

The third point is important as the successful completion of a platinum standard evaluation requires specialist and technical knowledge that may be unavailable within AusAID on a regular basis over the life of the project. As such, for selected projects, AusAID should consider outsourcing evaluations to universities or institutions focused on such designs to limit risk of non-completion. Clearly this approach should be cost-effective and as such dependent on the overall project cost, the possible learning outcomes and relevance to future projects. With an aid program valued at over 4 billion dollars, surely there is more than just one such project as a candidate for platinum evaluation.

Why should AusAID take this approach?

1. AusAID has a goal of being the best donor possible. This cannot be accomplished without appropriate assessment of aid effectiveness. To accomplish this, undertaking ex-post only evaluations will not suffice.
2. Taxpayers deserve to have the most rigorous understanding of 'bang for their buck'.
3. Without the existence of appropriate counterfactuals, project impact may be overstated.
4. Platinum standard evaluations can be presented as 'flagship' studies by AusAID.
5. Influential parliamentarians are increasingly asking for the use of randomized trials (see this [post](#) by Dr. Andrew Leigh, MP).

Assuming technical expertise exists, a key difficulty is convincing partners to undertake such platinum standard evaluations. Partner concerns may be warranted; for example, it may be impractical to administer an experimental design. More often than not, partners do not believe the value of such rigorous evaluations outweigh the costs of potential negative results and the additional administrative burdens they may impose. Moreover, evaluators are fighting against strong incentives from project managers against undertaking rigorous evaluations.

Without at least some mixed method evaluations incorporating experimental designs, it is difficult to argue causal impact at a high level of credibility. AusAID, by holding the purse strings, can request that such evaluations are administered (where practical) across specific projects. Similarly, for organisations receiving core AusAID funds, AusAID can request that a certain number of such evaluations are undertaken. The numbers promised and delivered can be tracked on a regular basis.

To be clear, platinum style evaluations which incorporate various qualitative techniques and a randomized quantitative component are difficult to undertake and are impractical in the majority of cases. But they are possible. At a minimum, AusAID should consider institutionalising platinum evaluations into a small percentage of its programs.

*Dinuk Jayasuriya is a Post-Doctoral Fellow at the [Development Policy Centre](#), Australian National University. Previously, he was a Monitoring and Evaluation Officer with the World Bank Group.*