

Performance benchmarks submission

Robin Davies and Stephen Howes¹

Contents

| | |
|--|----|
| 1. Introduction and summary of recommendations | 2 |
| 2. Historical overview | 3 |
| 3. Performance benchmarks..... | 7 |
| 3.1. Clarify consequences | 7 |
| 3.2. Stick to what we can control | 8 |
| 3.3. Give due weight to process | 8 |
| 3.4. Avoid perverse incentives | 9 |
| 3.5. Stay above the detail: benchmark systems..... | 10 |
| 3.6. Use external data and transparent methods..... | 10 |
| 3.7. Embed the performance framework in a three-tier system, where the third tier has a strong focus on aid processes..... | 11 |

¹ Robin Davies is the Associate Director and Professor Stephen Howes the Director of the Development Policy Centre of the Crawford School of Public Policy at the Australian National University.

1. Introduction and summary of recommendations

The government has [announced](#) it wants to develop rigorous performance benchmarks for the aid program.

As well as releasing a [consultation paper](#), the government has outlined its initial thinking on how benchmarks might be applied through the [speech](#) of the Minister for Foreign Affairs at the Australian National University on 13 February 2013.

Benchmarking is one strategy to increase the performance orientation of the aid program. In this submission, we first put the current exercise in context by providing a critical historical overview of efforts to enhance the performance orientation of the aid program, going back a decade or so (Section 2). We then go on to make specific recommendations in relation to benchmarking (Section 3).

A one-sentence summary of our recommended approach would be: **implement and extend past aid reforms, and focus on benchmarking process.**

Our recommendations are as follows:

Recommendation 1. In the benchmarking process, the Government should aim to improve the implementation of and extend the substantial aid reforms of the past decade.

Recommendation 2. The Government should commission a review by the Office of Development Effectiveness of efforts over the last ten years to link aid funding to performance and results.

Recommendation 3. Performance benchmarks should be used for redistributive purposes, not to determine aggregate aid levels.

Recommendation 4. Country and other (e.g. multilateral) allocations should certainly be influenced by performance considerations, but should not be tied rigidly to the achievement of policy or governance reforms that are beyond Australia's control.

Recommendation 5. Benchmarking the efficiency and effectiveness of aid management processes should be accorded at least as much importance as tracking policy-dependent benchmarks or targets that relate to aid outputs or outcomes.

Recommendation 6. Handle with caution and preferably avoid input-based benchmarks such as sectoral spending commitments, ensuring that, where used at all, they reflect a bottom-up consideration of what is needed and feasible.

Recommendation 7. Benchmark the use of key aid management systems by committing that they will be fit for purpose and consistently applied, and indicate how their fitness and use will be monitored.

Recommendation 8. The data and assumptions underlying the choice of benchmarks, and the data on which assessments of performance against benchmarks are based, should be reported in detail.

Recommendation 9. Externally verified performance data should be used wherever possible. Where this is not possible, internally generated data should be subject to random audit by the Office of Development Effectiveness or outside parties.

Recommendation 10. The aid program should conduct regular aid stakeholder surveys and use the resulting information, particularly trend information, as an input into the whole-of-program performance assessment process. Consideration could be given to defining one or more qualitative benchmarks relating to stakeholder perceptions.

Recommendation 11. Performance benchmarks should be consistently used in all aid performance reporting. They should be defined within the three-tier framework already adopted for the 2012 Comprehensive Aid Policy Framework, and should give due weight to process benchmarks vis-à-vis 'headline' policy-related or 'results' benchmarks.

2. Historical overview

One of the main recommendations of the 2006 aid [White Paper](#), which was commissioned by the then Minister for Foreign Affairs, Alexander Downer, was that the performance orientation of the aid program should be strengthened at all levels. There was an explicit statement that the aid scale-up to which Prime Minister Howard had committed in 2005² would be subject to improved aid effectiveness (e.g. p. ix). According to the White Paper, this would be done in four ways.

- Strengthen country strategies to include a performance framework.
- Strengthen the link between performance and allocations, between and within countries, including through the provision of performance incentives, which would constitute up to 10 per cent of the aid program.
- Build performance into individual activities by enhancing approaches that link payments to deliverables.
- Create the Office of Development Effectiveness, and strengthen the performance management system of the aid program.

Following the release of the White Paper, a lot of work was done to implement these reforms. Concrete measures included:

- The Office of Development Effectiveness (ODE) was created.
- The aid performance management system was strengthened.
- A \$116 million Performance Incentive Fund was created to provide additional funds to well-performing country programs.
- Australia joined the Global Partnership for Output-based Aid and started trying to link aid to outputs in some of its own activities.

² Mr Howard announced at the September 2005 UN World Summit that Australia would 'double its overseas aid allocation from 2004 levels to around \$4 billion by 2010.'

But there were also limitations:

- Country strategies became more elaborate, but not necessarily more effective.
- Although the Performance Incentive Fund was created, allocating the funding was tough going, the funds took a long time to spend and the Fund was never scaled up.
- It proved difficult to embed performance as a factor guiding project and country allocations.
- The Government of the day seemed unwilling to publish critical reports by ODE.

Overall, there was definite progress but the task of improving the performance orientation of the aid program was more difficult than had been anticipated. One factor here was that when Labor came to power at the end of 2007 they stopped using the White Paper as a point of reference and failed, for a very considerable period of time, to put in place any other strategy. As a result there was a lot of drift and fragmentation.

The next big effort at aid reform came through the [Hollway Review](#) of 2011 – the *Independent Review of Aid Effectiveness*.³ This also had a strong focus on performance, which manifested itself in six main recommendations, summarised below.

1. The expansion of the aid program – which was then well underway – should be based on performance. Better performing country programs should be scaled up more. PNG and most of the Pacific should receive smaller increases, and Asia larger increases.
2. The Government should move toward performance-based multilateral funding by undertaking an explicit ranking of the performance of multilateral funds and organisations. (The review panel thought this would be much harder to do with NGOs.)
3. More should be done to embed performance into individual projects and programs.
4. The Office of Development Effectiveness should be strengthened by having it placed under the technical oversight of an Independent Evaluation Committee.
5. An overall performance framework for the aid program should be introduced, filling a substantial gap. Following the approach that is now becoming standard for multilateral development banks, this would define performance measures (i.e. benchmarks) at three levels or ‘tiers’: (i) the overall economic and social development of partner countries, (ii) the outputs and outcomes of Australia’s aid, and (iii) the effectiveness and efficiency of Australia’s aid administration. Performance would be reported in an Annual Review of Aid Effectiveness.
6. The scale-up of aid should be made subject to ‘hurdles’. This was the final recommendation of the report, recommendation 39. The hurdles were to

³ One of the authors of this submission, Stephen Howes, was a member of the review panel.

be of two broad types, relating to outputs/outcomes and structures/processes. The government would need to be satisfied, principally on the basis of a consideration of the Annual Review of Aid Effectiveness, that the aid program was generally on target to achieve agreed outputs and outcomes. And the government would need to be satisfied that the aid management reforms flowing from the Review were being satisfactorily implemented, so that the scale-up could be continued with confidence.

The government of the day accepted recommendations in all the above areas, but implementation of some of them was lacking. The Coalition was especially scathing, and in our view with some justification, about the lack of a serious response to recommendation 39, relating to the establishment of hurdles.

What lessons can the present government can draw from this experience?

First, the government has more than eight years' worth of work to draw on. Although neither the White Paper nor the Hollway Review use the term 'benchmarks,' they cover the same ground. The White Paper spoke in terms of increasing the performance orientation of the aid program. The Hollway Review spoke in terms of managing for results. The performance framework described in the government's consultation paper corresponds to what the Hollway Review called a 'summary scorecard' (though the previous government chose not to utilise exactly that kind of a tool). Contrary to what is implied in the consultation paper, there is already a set of performance benchmarks for the aid program. It can be improved on, but it exists.

Second, while the government should not overlook past reforms, it can certainly improve upon and better implement them. In the Development Policy Centre's [2013 Aid Stakeholder Survey](#), most respondents thought that the previous government's aid effectiveness strategy was appropriate, but far fewer were satisfied with its implementation. One good example of this is the Annual Review of Aid Effectiveness. The one and only such [review](#) released by the previous government was anodyne and uninformative. It lacked a rigorous evidence base and contained no critical assessment, particularly at the level of major country programs and partnerships. (The [second](#) such review, just released, does a somewhat better job – though we have not yet had an opportunity to consider it carefully – but is still working with a deficient results framework.) Another example is the Australian Multilateral Assessment. It has been completed, but there is no evidence that it has influenced any multilateral funding decisions to date.

Third, benchmarks should not be looked at in isolation, but as one part of a broader performance management system which ensures that everything 'under the bonnet' is present, in good working order and used when it should be used. The aid program has at least seven major management systems in play, aside from the overall results-setting, monitoring and reporting system itself:

- i. developing, reviewing and modifying **country strategies**;
- ii. assuring **individual activity quality** at the point of entry to the portfolio and during implementation;

- iii. undertaking post-completion **impact evaluation** of activities, clusters of activities and country strategies;
- iv. undertaking **partner performance assessments** of implementing partners and agents (multilateral organisations, NGOs, contractors and, on occasion, partner governments);
- v. achieving **value for money** through management efficiency measures, due diligence processes and the competitive, transparent selection of implementing agents;
- vi. **managing fiduciary risk**; and
- vii. developing and maintaining the **professional capacity** of public sector aid policy advisers, sector specialists, process specialists and general administrators.

Failure of any of the above systems will undermine the performance of the aid program, regardless of the benchmarks in place. We have made broader recommendations about measures that need to be taken to improve the performance of the aid program in our recent [submission](#) to the Senate inquiry into Australia's overseas aid program. We will not repeat them here, except to underline the importance of transparency. If there is a fall-off in transparency then gains from putting in place new and better benchmarks are unlikely to be realised.

Fourth, even if a comprehensive approach is put in place to improve the effectiveness of Australia's aid, we will still struggle to achieve effectiveness at certain times and in certain places. The most critical determinant of aid effectiveness is the performance of the recipient government ([Howes, 2014](#)). As long as our aid is concentrated in the Pacific, effectiveness will be a challenge.

Fifth, and related to the previous point, the experience of past efforts teaches us just how difficult is the enterprise of linking aid to performance. There are all sorts of things which make it difficult to have a performance orientation. It is often very difficult to measure aid success and attribute it to specific aid interventions. And then, even if we know if we are succeeding or failing, it can be difficult to act on that information, for example, for diplomatic reasons. This is not a counsel of despair, but we should be realistic, and learn from the lessons of success. A review by ODE of the success or otherwise of past attempts to link funding to results would therefore be very useful. A limited but relevant [review](#), which focused specifically on the use of performance incentive payments, was undertaken in 2003 by the former AusAID Office of Review and Evaluation, but a lot has happened since that time. There have been many relevant initiatives over the past decade, including the Pacific Partnerships for Development, budget support programs, the Performance Incentives Fund, the Papua New Guinea-specific Incentive Fund and output-based aid. The government should seek to learn lessons from these, and from the 2003 review.

| |
|--|
| <p>Recommendation 1. In the benchmarking process, the Government should aim to improve the implementation of and extend the substantial aid reforms of the past decade.</p> |
|--|

Recommendation 2. The Government should commission a review by the Office of Development Effectiveness of efforts over the last ten years to link aid funding to performance and results.

3. Performance benchmarks

We now turn to the more specific issue at hand, namely the introduction of (improved) performance benchmarks for the aid program. We have five general suggestions in this regard, one in each of the next five sub-sections. (In what follows, we are drawing heavily on the final part of our earlier submission to the Senate inquiry into Australia's overseas aid program.)

3.1. Clarify consequences

The *Independent Review of Aid Effectiveness's* recommendation 39, concerning 'hurdles' was premised on the availability of large annual aid increases until 2015-16. The consequence of failing to meet hurdles was to be a reduction in the rate of growth of the aid program as a whole. The situation is now very different: there will be no growth in the program in real terms for several years. The government therefore needs to clarify whether its \$5 billion aid volume commitment is conditional on the achievement of hurdles or benchmarks, such that aid might be reduced below \$5 billion if certain of them are not met. Such a scenario of reduction is in fact contemplated in the government's discussion paper on benchmarks.⁴

While it could, perhaps, be argued that the rate at which aid was scaled-up should be linked to the ability of the program to meet ever-more demanding benchmarks, there is no case for using benchmarks to influence total aid when aid is not increasing. This creates an unwelcome unpredictability. It would also make us unique among aid donors: no other donor is known to have threatened to abandon a volume commitment if it fails to achieve certain outcomes or meet certain aid management benchmarks. But the most fundamental point is that we already have a good aid program where, on average and at the margin, benefits are much higher than costs.

What would be more productive, and what is certainly much more common, is the use, also flagged in the government's discussion paper, of performance information for redistributive purposes, usually within broad categories of partnership. For example, performance information might not determine a donor's total allocation to multilateral organisations or NGOs but it might inform decisions about the allocation of resources among multilateral organisations or NGOs. This approach has the obvious advantage that it creates a climate of competition among peers, whether they be countries, international organisations or NGOs. In short, it is important to be clear about what are the consequences, and within what domain, of failing to meet hurdles or benchmarks.

Recommendation 3. Performance benchmarks should be used for redistributive purposes, not to determine aggregate aid levels.

⁴ See paragraph 4, p. 2, of the [consultation document](#).

3.2. Stick to what we can control

Benchmarks for aid quality and impact are one thing. Benchmarks for partner governments' own policies and governance practices are quite another. The government's discussion paper on benchmarks distinguishes these things but still brackets them together. In general there is no reason to think that varying countries' aid allocations up or down can, by itself, have any lasting influence on their governments' policies or governance practices. These things are essentially beyond Australia's control. This is why the standard approach to articulating results frameworks contains a top tier of development results for which donor accountability is not accepted.

However, where poor policies or practices degrade the effectiveness of jointly agreed activities, there is clearly a case for aid to be reallocated either to other purposes within the country—possibly purposes less reliant on partnership with the government—or to another country. Thus benchmarks for partners should relate to second-tier and third-tier results, not first-tier results. Broad aid 'compacts', involving high-level commitments on both the donor and partner government side and mutual accountability, might in some cases serve a useful purpose. But attempts should not be made to link the fulfillment of compact-style commitments by the developing country party to the overall level of aid provided by the donor party.

Recommendation 4. Country and other (e.g. multilateral) allocations should certainly be influenced by performance considerations, but should not be tied rigidly to the achievement of policy or governance reforms that are beyond Australia's control.

3.3. Give due weight to process

While output and outcome targets are useful both for aid management and for communications purposes, process-related measures are in fact more likely to yield information about the effectiveness of aid, and are of course necessary for assessments of efficiency. This is for several reasons, but the fundamental one is that it is intrinsically difficult, if not impossible, to assess aid's results in any rigorous way. Much better then, to monitor what one can observe, which are effective processes. One can be confident that, all other things being equal, effective processes will lead to effective aid. While this message can be misinterpreted as an expression of defeatism, to ignore it is to abandon control over aid's practical levers: its delivery mechanisms.

The importance of focusing on process is underlined by the fact that the whole 'aid effectiveness' [agenda](#) of the OECD Development Assistance Committee, while registering the importance of being guided by desired results, is dominated by process indicators. Given the difficulty of measuring development outcomes and attributing them to aid inputs, let alone the inputs of single donors, this is not surprising. In addition, process benchmarks are by nature more constant and readily monitorable and comparable over time, whereas outcome indicators tend to shift as geographical priorities and sectoral preoccupations shift. Thus, in any performance framework, it is important to delineate clearly a class of policy-

independent process-related benchmarks that yield information relevant for assessments of aid quality, alongside policy-dependent outcome indicators.

It is not our purpose here to propose specific benchmarks, but there are many specific areas in which process benchmarks might be expected to yield useful performance information about both DFAT and its partners and agents. A number of these are outlined below, with asterisks against those that did in fact figure to some extent—though often in quite a partial way—in the results framework set out in the 2012 Comprehensive Aid Policy Framework (CAPF).

- Degree of aid predictability (program-specific volume stability, and continuity in programming)
- *Degree of aid fragmentation (activity size)
- *Completeness and consistency in the application of quality assurance systems for activities at portfolio entry, in implementation and post-completion
- *Adequacy of portfolio (country and thematic) strategy development and review processes, including timeliness
- *Level of information transparency
- *Rigour of partner quality assessment processes (relating to partner government financial systems and to multilateral, NGO and contractor performance)
- Rigour and timeliness of annual reviews of aid effectiveness
- Appropriateness of administrative costs as a share of aid
- *Efficiency and openness of procurement, and adequacy of fiduciary risk management and loss recovery systems
- *Independence, quality and coverage of evaluations
- *Appropriate skills and gender mix among aid-focused staff
- *Adequate continuity of staffing in aid-related areas

Recommendation 5. Benchmarking the efficiency and effectiveness of aid management processes should be accorded at least as much importance as tracking policy-dependent benchmarks or targets that relate to aid outputs or outcomes.

3.4. Avoid perverse incentives

Input-based benchmarks can be quite problematic. For example, an arbitrary commitment to increase spending on X to a certain dollar amount or share of the program over a certain period of time might turn out to be unimplementable, or else might distort the shape of country programs and lead to the development of programs that do not have the full support of partner governments. Likewise, a commitment to increase the number of NGOs eligible for subsidies, or to send a certain number of volunteers overseas each year, can lead to suboptimal resource allocation. Benchmarks of this nature are sometimes useful and appropriate, but only if constructed on the basis of a consideration of what is actually needed and possible—that is, if they are constructed in a bottom-up rather than top-down fashion. It is tempting but not desirable to drive implementation of new policy priorities by setting top-down targets. A better approach is to pursue those priorities by making modest levels of additional

funding available to receptive country programs for pilot activities, or reallocating existing country program resources in cases where partner governments are particularly amenable. High-level targets can then be estimated on the basis of the projected results of activities for which there is known demand and feasibility.

Other cases where benchmarks should be avoided because of the likelihood that they would give rise to perverse incentives might include the following.

- *Arbitrarily determined process targets achieved (e.g. ‘at least 50 accredited Australian international development NGOs participating in ANCP by 2015–16’—an actual CAPF target)
- Certain levels of joint programming with other donors, or ‘trilateral’ cooperation with other donors and partners countries, achieved.

Recommendation 6. Handle with caution and preferably avoid input-based benchmarks such as sectoral spending commitments, ensuring that, where used at all, they reflect a bottom-up consideration of what is needed and feasible.

3.5. Stay above the detail: benchmark systems

Benchmarks can obviously operate at many levels, and they need to. Each individual activity should have clearly articulated goals together with strategies for their achievement. But at the level of the aid program as a whole, the positive impact of benchmarks will fall away rapidly as their number and complexity increases. A sensible approach here is, wherever possible, to focus benchmarks at the ‘systems’ level. As noted in the previous section, there are at least some seven systems which are important for aid management. In each case, rather than setting a series of specific, fine-grained benchmarks, a better approach might be to commit to the consistent application of fit-for-purpose systems, with such commitments tested regularly through independent processes, including ANAO performance audits and ODE’s own annual review of performance systems, which should be continued.

Consideration should be given to setting broad benchmarks at the level of DFAT’s main aid management systems—essentially giving a commitment that each system be consistently used, either as it is or following reform, and stating what accountability processes will apply—rather than seeking to define headline benchmarks for each specific point (including some of those listed just above) on which performance information might be desired.

Recommendation 7. Benchmark the use of key aid management systems by committing that they will be fit for purpose and consistently applied, and indicate how their fitness and use will be monitored.

3.6. Use external data and transparent methods

Wherever possible, benchmarks should be established on the basis of objective and public data. Unlike practice to date in Australia, but as observed elsewhere, the basis of calculations and the mix of data and assumptions that goes into them should be made transparent, both in connection with the setting of benchmarks

and the measurement of performance against them. This is perhaps where there is the biggest scope for improvement relative to the performance framework introduced under Labor, where data and assumptions were never clarified or made public.

It must be recognised that any benchmarking system gives rise to strong incentives to meet targets, and to over-emphasise the positive. Externally-collected data should be used wherever possible to counter at least the latter bias. Where reliance is placed on internal data, such data should be subject to random audits, by the Office of Development Effectiveness or other bodies.

The aid program is funded but not, for the most part, implemented by the Australian government. The many other parties responsible for the delivery of the aid program – NGOs, contractors, multilaterals, partner governments themselves – are very well informed about its strengths and weaknesses. The perceptions of key aid stakeholders should be considered in assessing progress against benchmarks, and explicit reference to stakeholder perceptions might also reasonably figure in one or more qualitative benchmarks. The aid program should regularly carry out aid stakeholder surveys, building on the one carried out by the Development Policy Centre in 2013, and analysed [here](#).

Recommendation 8. The data and assumptions underlying the choice of benchmarks, and the data on which assessments of performance against benchmarks are based, should be reported in detail.

Recommendation 9. Externally verified performance data should be used wherever possible. Where this is not possible, internally generated data should be subject to random audit by the Office of Development Effectiveness or outside parties.

Recommendation 10. The aid program should conduct regular aid stakeholder surveys and use the resulting information, particularly trend information, as an input into the whole-of-program performance assessment process. Consideration could be given to defining one or more qualitative benchmarks relating to stakeholder perceptions.

3.7. Embed the performance framework in a three-tier system, where the third tier has a strong focus on aid processes.

However benchmarks are ultimately elaborated, it will be important that once determined they are used consistently in key policy and accountability documents, such as the Comprehensive Aid Policy Framework (CAPF) or equivalent, Annual Program Performance Reports or equivalent, the Annual Review of Aid Effectiveness or equivalent, the DFAT Annual Report and the annual aid budget statement (the 'Blue Book').

In her 13 February 2014 [speech](#) at the Australian National University, the Minister for Foreign Affairs outlined in some detail her approach to the proposed benchmarking of Australian aid. She indicated that there would be three levels of benchmarks. At the top level would be 'a small number of high level targets' to 'assess the entire aid program against key goals and priorities.' The second level

would be 'the level of individual programs.' The third level would be the individual activity level. Assessments at the second level, basically at the country level, would be used to allocate funds between countries. Assessments at the third level would be used to shift funding towards better performing activities.

Superficially, this sounds similar to the three-tier results framework put in place by the previous government in the 2012 CAPF, but the only similarity is actually the number of tiers. As noted in section 1, the CAPF's first tier was partner-country performance against the MDGs, the second tier Australian aid's performance in achieving aggregated results (such as schools built), and the third tier the aid administration's performance against indicators of organisational and operational effectiveness.

All three tiers might, it appears, be different under the Coalition. We have no objection to the first tier being changed. We all know which countries are on or off track against the MDGs. It might well be more useful to have as our first tier, as the Minister has suggested, a small number of key government commitments, such as whether we are indeed focusing on the Asia-Pacific, or effectively engaging with the private sector to improve development outcomes.

In our [critique](#) of the [2011-12 Annual Review of Aid Effectiveness](#) produced by AusAID we criticised the absence of attention to the performance of country programs and projects. Information on both of these fronts is assiduously collected by various aid program performance management systems, but this information was thoroughly neglected in the 2011-12 annual review, the only such review produced under the CAPF results framework. So we would also welcome greater attention to country and project performance, and the use of this information to reallocate money to better programs and projects.

Our concern, however, is that the three tiers outlined by the Minister seem to leave no place for the old third tier, thus squeezing out any notion that key aid management processes will be benchmarked. Yet, as we have argued, it is more important to have benchmarks at the process level than at any other.

| |
|---|
| <p>Recommendation 11. Performance benchmarks should be consistently used in all aid performance reporting. They should be defined within the three-tier framework already adopted for the 2012 Comprehensive Aid Policy Framework, and should give due weight to process benchmarks vis-à-vis 'headline' policy-related or 'results' benchmarks.</p> |
|---|