



Australian Government

Department of Foreign Affairs and Trade

2014 Quality Review of Aid Program Performance Reports

Office of Development Effectiveness

April 2015

© Commonwealth of Australia 2015

ISBN 978-0-9874848-5-7

Published by the Department of Foreign Affairs and Trade, Canberra, 2015.

Disclaimer: The views contained in this report do not necessarily represent those of the Australian Government.

With the exception of the Commonwealth Coat of Arms and where otherwise noted all material presented in this document is provided under a Creative Commons Attribution 3.0 Australia (<http://creativecommons.org/licenses/by/3.0/au/>) licence. The details of the relevant licence conditions are available on the Creative Commons website (accessible using the links provided) as is the full legal code for the CC BY 3.0 AU licence (<http://creativecommons.org/licenses/by/3.0/au/legalcode>). The document must be attributed as *Office of Development Effectiveness (2015). 2014 Quality Review of Aid Program Performance Reports*, ODE, Australian Government Department of Foreign Affairs and Trade, Canberra.

Office of Development Effectiveness

The Office of Development Effectiveness (ODE) is an independent branch within the Australian Government Department of Foreign Affairs and Trade (DFAT). ODE monitors the Australian aid program's performance, evaluates its impact, and contributes to international evidence and debate about aid and development effectiveness. ODE's work is overseen by the Independent Evaluation Committee (IEC), an advisory body that provides independent expert advice on ODE's evaluation strategy, work plan, analysis and reports.

www.ode.dfat.gov.au

Contents

Acknowledgements	1
Executive summary	2
Findings	2
Conclusion and recommendation	5
DFAT response to recommendation	6
1 Introduction	7
1.1 Aid Program Performance Reports	7
1.2 Aid performance management in DFAT	7
1.3 Country and regional program performance management	8
1.4 Methodology	10
2 Quality review	12
2.1 Context and strategic orientation	12
2.2 Progress assessment	14
2.3 Partner contribution	17
2.4 Management consequences	19
2.5 Other observations	20
3 Performance benchmarks	21
3.1 Performance benchmark characteristics	21
3.2 Performance benchmarks as measures of performance	25
4 How well do APPRs reflect changes in aid program performance?	26
Appendices	29
Appendix A: Terms of reference (excerpt)	29
Appendix B: APPR quality review assessment questions	30
Appendix C: APPR quality by size and whether program has a performance assessment framework	32
Appendix D: Cornerstone question assessment comparison over two years	33
Appendix E: Summary tables of quality ratings against all cornerstone questions and subquestions	34
Abbreviations	38

Acknowledgements

The Office of Development Effectiveness (ODE) thanks all those who contributed to this report. The 2014 Quality Review of Aid Program Performance Reports (APPRs) was undertaken for ODE by Jo Hall and Kari Sann from IOD PARC with assistance from Julian Gayfer from IOD PARC who led last year's APPR review. Serin Werner from ODE provided day-to-day guidance on the review process and together with Jo Hall and Kari Sann, authored this report. The report was overseen by David Slattery from ODE.

Thank you to those who provided comments on drafts of the report, including the ODE Independent Evaluation Committee; Tom Davis from DFAT's South and West Asia Division; and Ian Kidd, Andreas Zurbrugg and Andrew Collins from DFAT's Aid Management and Performance Branch.

The analysis, conclusions and recommendations in the quality review are those of ODE.

Executive summary

Each year since 2008, country and regional programs have reported progress against their objectives in annual Aid Program Performance Reports (APPRs). APPRs are published on the Department of Foreign Affairs and Trade (DFAT) website, and serve three main purposes: strengthening program management and effectiveness; providing accountability; and facilitating policy dialogue with development partners.

While the Australian aid performance management system has changed significantly in recent times—including the release of a new aid policy* and an accompanying performance framework†—APPRs remain an integral performance management and reporting tool for the Australian aid program.

The Office of Development Effectiveness (ODE), under the guidance of the Independent Evaluation Committee, provides independent oversight of departmental performance reporting. As part of this role, ODE has carried out quality reviews of APPRs every year since they were introduced. These examine the quality of APPRs and highlight strengths and opportunities for improvement. This year's quality review also looked at APPRs' treatment of performance benchmarks and mutual obligations, which are new requirements.

The 2014 ODE Quality Review of APPRs assessed the quality of 28 2013–14 APPRs.‡ Key findings, conclusions and recommendations follow.

Findings

APPR quality is consistent with previous year

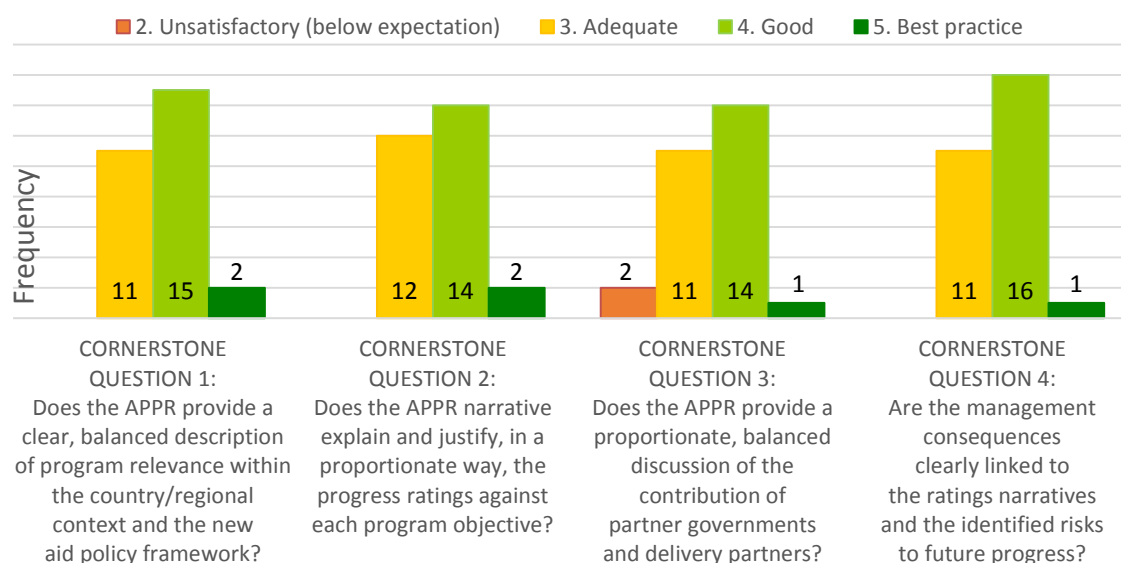
The quality of 2013–14 APPRs is consistent with the previous year's: they are well written, informative descriptions of program activities and progress towards objectives. They discuss program successes and failures in a manner that is largely evidence-based, and balanced. Almost all APPRs were rated as either good or adequate against all four cornerstone questions, which assess key elements of APPR quality (Figure 1).

* *Australian aid: promoting prosperity, reducing poverty and enhancing stability*, DFAT, Canberra, 2014.

† *Making Performance Count: enhancing the accountability and effectiveness of Australian aid*, DFAT, Canberra, 2014.

‡ The Nauru APPR was the 29th and final APPR for 2013–14. It was submitted too late for inclusion in the quality review.

Figure 1 Almost all APPRs were rated as “good” or “adequate” against each cornerstone question



APPRs identify development challenges and demonstrate program alignment with new policy directions, but explaining program strategy is a weakness

All APPRs do a good job of describing the development challenges of the country/region. Relatively few, however, provide good descriptions of program strategy (what Australian aid will do and the logic of how this will contribute to achieving program objectives). This is a fundamental feature of reports, as the essence of the APPRs is to reflect on how well this strategy is working. Although there are examples of good practice, this remains a weakness. DFAT’s Aid Management and Contracting Division (ACD) advise that Aid Investment Plans*, due in September 2015, should assist with the articulation of program strategy.

The vast majority of APPRs pay particular attention to the new aid policy priorities of aid for trade, engaging with the private sector and, in particular, empowering women and girls. While integrating gender equality into all investments is still a work in progress for the aid program,† the attention it was accorded in 2013–14 APPRs demonstrates this issue is being appropriately prioritised.

The quality of partner assessments is mixed

The performance of partner governments is adequately discussed throughout most reports (including in the new ‘mutual obligations’ section, addressed below). Whole-of-government partners are mentioned where appropriate, but their performance is not assessed. Performance assessment of multilateral and non-government organisation (NGO) partners is mixed, with some APPRs making detailed assessments and others only mentioning them as partners.

* Aid Investment Plans will set out the direction for a country or regional program and link objectives, aid programming and results.

† In 2014 Aid Quality Checks, 74 per cent of investments were rated as performing satisfactorily or above on gender equality.

Almost all APPRs provide specific and actionable management responses

This year's management responses may be slightly less concrete than in 2012–13 APPRs. However, it is perhaps unsurprising that with the budget uncertainly prevailing at the time they were authored, some APPRs are less concrete about management actions than they have been in previous years. This is because a large proportion of management responses have budget implications.

APPRs for smaller programs tend to be of poorer quality

As was the case for the previous year's APPRs, we found that reports for programs valued at less than \$50 million tend to be lower quality than those for larger programs. This is likely to be due to smaller programs having fewer resources to develop APPRs and underlying performance frameworks.

Most APPRs provide a frank and balanced assessment of performance against their objectives; however, some objectives are not clearly framed

Programs rate progress towards achieving objectives in APPRs. All APPRs are at least satisfactory in terms of explaining and justifying ratings given against objectives. Eight of the 28 APPRs, however, had one or more objective that was too broad or vague, and it was not clear how performance ratings had been derived. If ratings against objectives are to be used as indicators of program performance, then at minimum program objectives should be clear about what is trying to be achieved and the basis on which performance is judged.

If ratings against APPR objectives are used to measure progress over time, there needs to be more stability of objectives and how progress against them is measured

Ratings against objectives are also commonly used by the department to indicate changes in aid program performance over time. It is important to understand the limitations of using these ratings in this way—especially given the strengthened policy emphasis on linking performance with budget allocations. Comparative analysis of the objectives and ratings in 2012–13 and 2013–14 APPRs shows a good deal of change in the actual objectives, and/or the way progress against them is measured, over the two years, which impacts on their comparability.

New elements of APPRs will take time to bed down

The introduction of the new performance framework for the Australian aid program added two new elements to APPRs: *performance benchmarks* and *mutual obligations*. Both will take time to bed down. The variable quality of the initial performance benchmarks 'piloted' in 2013–14 APPRs shows that work is needed for them to become more robust indicators of program performance. With improvement, they could be considered alongside the traffic light ratings in APPRs to help inform assessments of the relative performance of country and regional programs. The new mutual obligations section of APPRs is interpreted differently in different APPRs, which may reflect a lack of clarity in the current APPR guidance on how to approach this requirement.

Performance assessment frameworks are important

Performance assessment frameworks (PAFs) articulate the logic for how aid interventions will achieve their objectives. They align program objectives with outcomes, indicators and baselines, and thus facilitate logical program planning, and enable tracking of progress toward objectives. Reinforcing the findings from last year's APPR quality review, we found that programs with good

quality PAFs are generally more able to credibly report on performance—this includes being better able to set measurable objectives, where it is clear what is trying to be achieved and the basis on which performance is judged; justify ratings against objectives; and identify better sets of performance benchmarks.

Conclusion and recommendation

The vast majority of the 28 APPRs reviewed are satisfactory, good or best practice in the assessment of key quality features, with only two APPRs falling below standard on the discussion of partner contribution. Reports are generally well written, highly informative, and appear to be used as performance management tools. While a couple of reports paint an overly rosy picture, most provide a frank and balanced assessment of performance. The APPRs of smaller programs tend to be of lower quality than those of larger programs. This is likely to be due to smaller programs having fewer resources to develop APPRs and underlying PAFs.

A highlight in this year's APPRs was strong discussion around gender equality issues in most reports.

Two persistent issues face some APPRs. The first is that many APPRs do not clearly explain program strategy. Aid Investment Plans, due in September 2015, should assist with the articulation of program strategy. The second is that too many of the objectives in APPRs lack clarity. Eight out of the 28 APPRs had at least one objective that was too broad or vague, and/or lacked transparent indicators of progress. Without objectives that are clear and measurable, it is difficult to explain and justify ratings of progress against program objectives. It also limits the capacity of APPRs to provide meaningful insight into the relative performance of country and regional programs.

Any assessments of year-on-year changes in program performance based on APPR ratings should take into account changes in APPR objectives, and the way progress against them is measured. Our comparative analysis of the objectives and ratings in 2012–13 and 2013–14 APPRs demonstrates that there has been significant movement in the objectives of some programs between the two years. Programs change, and get better at assessing performance; hence, some changes to objectives and their indicators are inevitable. Nonetheless, if the department wishes to strengthen the capacity of APPRs to measure performance over time, country and regional programs should—to the extent possible—ensure stability of objectives, and indicators from year to year (including transparency of how progress against them is measured).

The new aid performance framework requires that country and regional programs set performance benchmarks to assess progress toward objectives. The 2013–14 APPRs included 'pilot' benchmarks to assess performance in 2014–15. DFAT intends to use measures of progress against performance benchmarks to inform annual budget allocations, so it is important that they are robust performance indicators. ODE considers that with improvement, the performance benchmarks can play a role in assessing program performance—but that the benchmarks alone will not enable the most credible assessment. In ODE's view, the benchmarks should be used in combination with other measures, especially APPR 'traffic light' ratings, to inform assessments of program performance.

Both this and last year's quality reviews of APPRs demonstrate that the credibility of APPRs' assessment of program performance is enhanced when programs can draw on good quality PAFs. Currently PAFs are not required for programs valued at less than \$50 million. ODE considers PAFs should be required for all programs that produce APPRs.

Recommendation

All country and regional programs that produce APPRs should have, and be supported to develop, high-quality performance assessment frameworks as part of their Aid Investment Plans. Program objectives, performance benchmarks and indicators should all be drawn from these frameworks.

DFAT response to recommendation

Noted. DFAT recognises the importance of APPRs providing evidence-based assessments of program performance against AIP objectives. AIP and APPR processes have been streamlined since the 2014 Quality Review of APPRs was undertaken. A more proportional approach is now being applied based on the size and risk profile of the country or region. Performance Assessment Frameworks can be a useful management tool contributing to the identification of performance benchmarks and are a mandatory requirement for programs over \$50 million.

1 Introduction

1.1 Aid Program Performance Reports

Aid Program Performance Reports (APPRs) are a critical part of DFAT's aid performance assessment system. Introduced in 2007, they annually assess the performance of country and regional programs against their objectives. In this way, they help the aid program to track its contribution to country and regional-level changes. APPRs are the only mechanism that take a whole-of-aid view of bilateral and regional programs (i.e. Official Development Assistance delivered by all Australian Government departments, not just DFAT).

APPRs serve three main purposes: strengthening program management and effectiveness; ensuring accountability; and facilitating dialogue with development partners.

APPRs provide an opportunity for programs to not only reflect on how they are progressing towards their objectives, but also have their assessment of program performance tested by peers. Sector specialists, program and quality managers and advisers, joined on some occasions by representatives from whole-of-government partners, read the draft APPRs and provide written and verbal feedback through the peer review process. APPRs are published and so contribute to the transparency and accountability of Australia's aid program.

Finally, APPRs are a key reporting link between Aid Quality Checks (until recently these were named Quality at Implementation Reports), which report at the investment (project) level, and whole-of-aid program reporting. Drawing on the information contained in APPRs enables DFAT senior management and the Australian Government to understand how the aid program is delivering on its objectives. This can inform decisions on significant shifts in country or regional strategies and associated allocation of resources.

Role of the Office of Development Effectiveness in reviewing APPRs

The Office of Development Effectiveness (ODE) provides independent oversight of departmental performance reporting. As part of this role, ODE has carried out Quality reviews of APPRs every year since they were introduced in 2008. These reviews examine the quality of APPRs and highlight strengths and opportunities for improvement. ODE's analysis is submitted to the Independent Evaluation Committee for review, discussion and endorsement.

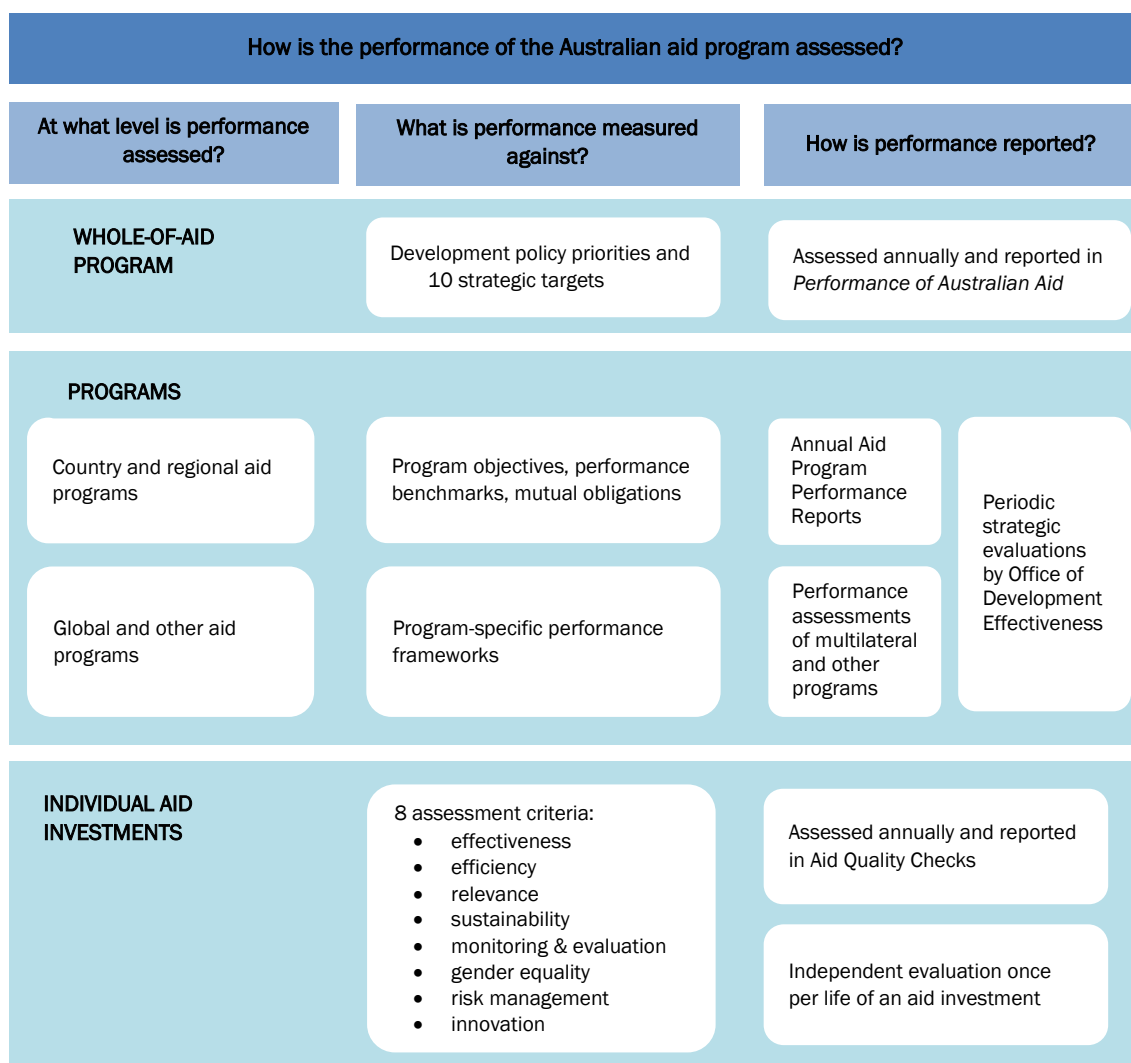
1.2 Aid performance management in DFAT

The Australian aid performance management system has changed significantly in recent times. This includes the release of a new aid policy* and an accompanying performance framework† in June 2014. The intent of the new framework, *Making Performance Count*, is to streamline aid performance management, and strengthen the focus on value for money as well as the link between performance and budget decisions.

* *Australian aid: promoting prosperity, reducing poverty and enhancing stability*, DFAT, Canberra, 2014.

† *Making Performance Count: enhancing the accountability and effectiveness of Australian aid*, DFAT, Canberra, 2014.

Figure 2 Performance assessment in the Australian aid program



Making Performance Count operates on three levels (Figure 2):

1. At the whole-of-aid program level: where 10 strategic targets focus the aid program to key overall goals and priorities.
2. At the country/regional program level: where annual APPRs report on performance against program objectives, performance benchmarks and mutual obligations.
3. At the investment level: where Aid Quality Checks annually assess the performance of individual investments against eight assessment criteria.

1.3 Country and regional program performance management

Planning country and regional programs

Engagement between the Australian aid program and aid recipients occurs mainly through country and regional programs. DFAT and partner governments agree on the objectives of these programs, which are pursued through a combination of direct investments (including aid

activities of other Australian Government departments), policy dialogue*, and the contribution of partners, including partner governments.

Aid Investment Plans, to be in place by September 2015, will set out the directions for country and regional programs (replacing Country Assistance Strategies), describing where, why and how Australian aid is delivered, generally over a four year duration in line with forward budget estimates. They will present a line of sight from Australia's national interests and policy priorities, as well as current and planned aid investments, to the results we expect to achieve in a country or region.

Performance benchmarks and mutual obligations

Making Performance Count specifies that each country and regional program must identify performance benchmarks to assess how well the program meets its Aid Investment Plan objectives. These are essentially targets that are, "tailored to the different development contexts and priorities of the countries we work in, making the way we measure performance more relevant to us and our partner governments." †

Aid Investment Plans will also outline the mutual obligations of the Australian and partner governments that are particularly important to the partner government's development priorities and Australia's aid investments.

Recent revisions to APPRs

The 2013–14 APPRs—on which this quality review focuses—were prepared according to the revised APPR templates and guidance notes that DFAT issued in 2014. New features include the requirements that programs must:

- › adhere to page limits that are relative to program budgets‡
- › describe mutual obligations with partner governments (and/or key regional organisations in the case of regional programs) that contribute to shared development goals, and what progress has been made in reaching any mutually agreed outcomes§
- › identify a small initial set of performance benchmarks to assess performance for 2014–15.**

It will take time for APPR guidance, templates and processes under the new performance framework—including the relationship between APPRs and Aid Investment Plans—to bed down and for staff to adjust to the new requirements. The period before the Aid Investment Plans are in place is an opportunity to phase in and learn from piloting the new ideas (performance benchmarks and mutual accountability).

* In its simplest form, aid policy dialogue is a conversation with policy-makers on the effectiveness of the overarching policy framework in which the aid is delivered. It provides an opportunity for stakeholders, including donors, to understand, inform and contribute to the development of that framework and the associated public expenditure choices. Policy dialogue is not just a technical challenge—the quality of the process also matters. Policy dialogue may, in practice, take many forms.

† *Making Performance Count: enhancing the accountability and effectiveness of Australian aid*, DFAT, Canberra, 2014, p.13.

‡ Programs with a total value of less than \$50m in 2013–14 were restricted to a maximum page length (excluding annexes and title pages) of 10 pages; \$50m–\$300m to 15 pages; and more than \$300m to 20 pages.

§ Mutual obligations will be more fully articulated in Aid Investment Plans.

** This only applies to 2013–14 APPRs; future APPRs will report on progress against the longer term performance benchmarks (typically covering a four-year period) that will be articulated in Aid Investment Plans.

1.4 Methodology

The 2014 Quality Review of APPRs has two objectives:

1. To report on the reasonableness of the performance narratives and the use of evidence.
2. To analyse the coverage, key characteristics and quality of the performance benchmarks proposed.

To do this, the review draws on several sources of evidence:

- › previous ODE APPR quality reviews
- › 28 of 29 2013–14 APPRs (Table 1)*
- › all 13 available performance assessment frameworks (PAFs; explained further in Section 2.2 below)
- › 2012–13 APPRs to compare performance and quality.

The review of APPR quality in Chapter 2 was guided by assessment criteria (Appendix B) comprised of 18 subquestions that assess key quality aspects of APPRs, and which in turn informed assessments/ratings against four cornerstone questions:

1. **Context and strategic orientation.** Does the APPR provide a clear, balanced description of program relevance within the country/regional context and the new aid policy framework?
2. **Progress assessment.** Does the APPR narrative explain and justify, in a proportionate way, the progress ratings against each program objective?
3. **Partner contribution.** Does the APPR provide a proportionate, balanced discussion of the contribution of partner governments and delivery partners?
4. **Management consequences.** Are the management consequences clearly linked to the ratings narratives and the identified risks to future progress?

Table 1 2013–14 APPRs assessed in this quality review

APPR type	Program
Country APPRs	Afghanistan, Bangladesh, Burma, Cambodia, Fiji, Indonesia, Kiribati, Laos, Mongolia, Nepal, Pakistan, Palestinian Territories, Papua New Guinea, Philippines, Samoa, Solomon Islands, Sri Lanka, Timor-Leste, Tonga, Tuvalu, Vanuatu, Vietnam
Regional APPRs	East Asia, Pacific, South Asia, Sub-Saharan Africa
Other APPRs*	Australian-NGO Cooperation Program (ANCP), Mekong Water Resources Program

* ANCP and Mekong Water Resources Program were not required to write APPRs but elected to do so voluntarily.

* The Nauru APPR was submitted too late for inclusion in the quality review.

Each APPR was reviewed and rated (with comments) against the same quality rating scale as last year’s APPR review (Figure 3). Compliance with DFAT’s current APPR guidance determined the difference between a rating of two (unsatisfactory) and three (adequate); that is, ***no APPR was rated less than adequate against any question if it complied with the guidance***. In this way, changes in APPR guidance since last year were taken into account. The ratings and comments were extensively moderated among the review team members and against last year’s assessments to maximise comparability. Ratings and patterns in performance were analysed by program size and location. Good practices as well as common weaknesses in all areas of APPR reporting were identified.

Figure 3 APPR quality rating scale

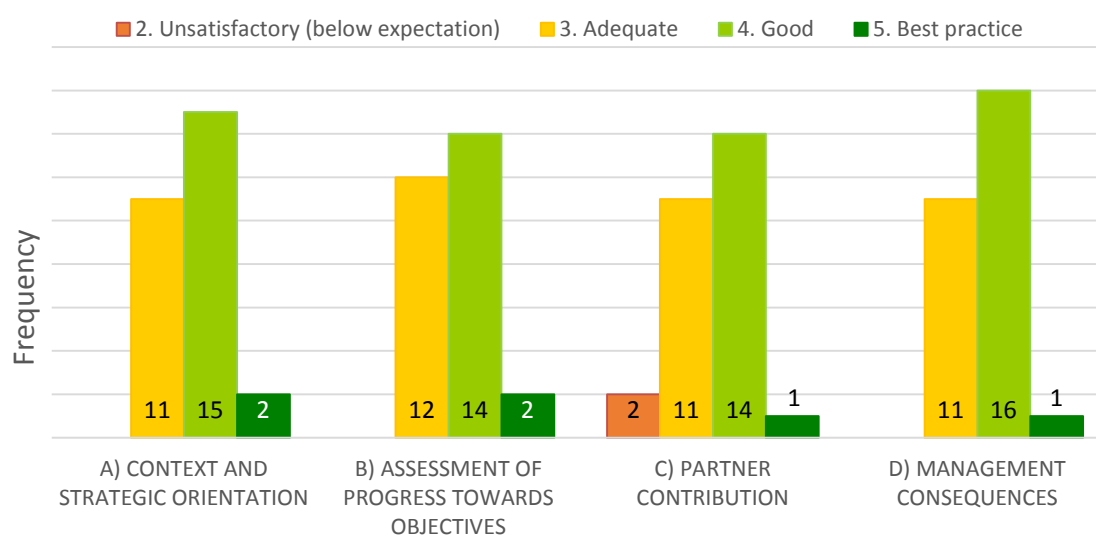
Rating	5	4	3	2	1
Explanation of rating	Example of best practice	Good	Adequate / Complies with guidelines	Unsatisfactory / Below expectations	Unsatisfactory / Weak

Note: Ratings of 2 or below indicate unsatisfactory quality. Ratings of 3 and above indicate satisfactory quality—a rating of at least 3 was given where the APPR complied with the APPR guidance. A rating of 4 indicates that additional information/analysis was given that enhanced the quality area. A rating of 5 reflects particularly high quality in the given area.

2 Quality review

The quality of 2013–14 APPRs is consistent with the previous year's: they are well written, informative descriptions of program activities and progress towards objectives that include evidence-based, balanced discussions of successes and failures. Almost all APPRs were rated as either good or adequate against all four cornerstone questions (Figure 4). Findings against each cornerstone question are discussed in turn below. Appendix E provides details of ratings against *all* cornerstone questions and subquestions.

Figure 4 Almost all APPRs were rated as “good” or “adequate” against each cornerstone question

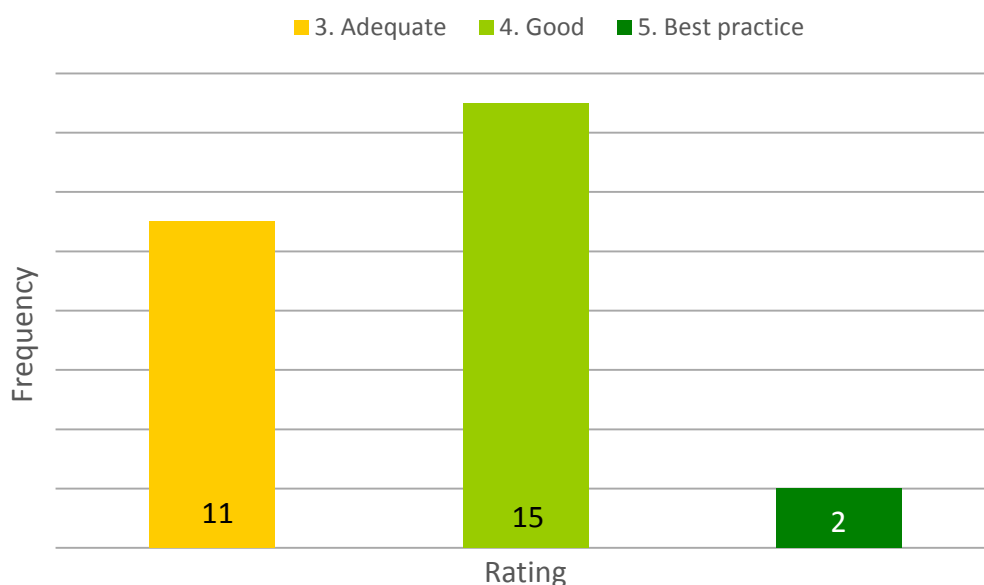


2.1 Context and strategic orientation

CORNERSTONE QUESTION 1	Does the APPR provide a clear, balanced description of program relevance within the country/regional context and the new aid policy framework?
Subquestion 1a	Does the APPR acknowledge, respond to and reflect the new aid policy?
Subquestion 1b	Does the context analysis explain the key development challenges/needs of the country/region, with an emphasis on changes that have occurred during the reporting period? Is this analysis used to explain the Australian program objectives?
Subquestion 1c	If there is a stated strategy or 'theory of change' does this explain how program objectives will be achieved, and does this explanation reference the country/regional context analysis? In the absence of a stated strategy or theory of change, is some other explanation for achieving program objectives, given the context, provided?
Subquestion 1d	Does the report make clear the proportion of Australia's aid in relation to other donors and national indicators, and the program and strategy implications of that?

This section looks at how successfully the reports describe the program context; reflect the logic of how Australian aid contributes to achieving program objectives; explain the size and significance of Australian aid in the context; and respond to Australia's aid policy.

Figure 5 Frequency of quality ratings against ‘context and strategic orientation’



Program areas are expected to position the country or regional program within the current context, and this requires adapting the program strategy from year to year as the context changes. All the APPRs do a good job of describing the development context of the country or region. Overall, all the APPRs are at least satisfactory in explaining the relevance of the program to the country or regional context (Figure 5). Better APPRs describe the position of Australian aid within that context by:

- › **Confirming objectives for Australian aid that make sense within the context**, and clarifying the sectors or areas of work and outlining what will be achieved in these areas.
- › **Reflecting how Australian aid contributes to achieving the strategy objectives**. Discussion of program logic or strategy, while less prominent than before as a requirement in the new APPR guidance*, is a fundamental feature of well written APPRs, because the essence of the APPR is to reflect on how well this strategy is working. Yet, clearly describing the program strategy remains a weakness in many APPRs. Good practice examples are in the APPRs for the Philippines (see Box 1), Burma and Pakistan.
- › **Describing the relative size and significance of Australian aid and the implications of this**. Most APPRs describe the size of Australia’s contribution relative to other donors. Better APPRs go on to explain the significance of Australian aid relative to the country’s own resources. Only a few APPRs, however, explain how the program has been designed to reflect the size and position of Australian aid. This positioning helps clarify the nature and emphasis of Australia’s efforts.
- › **Responding to the new aid policy**. The new aid policy strengthens the emphasis on private sector engagement, aid for trade, and gender equality and women’s empowerment. Some APPRs do not reflect the implications of the changed policy environment, while many are well ahead in adjusting their program to match. The Timor-Leste program is probably the most advanced.

* The new APPR template, under the “Context” section, prompts APPR authors to answer (among other questions): “Does the program/strategy remain relevant to the current context?”

Box 1 Good practice in strategic orientation

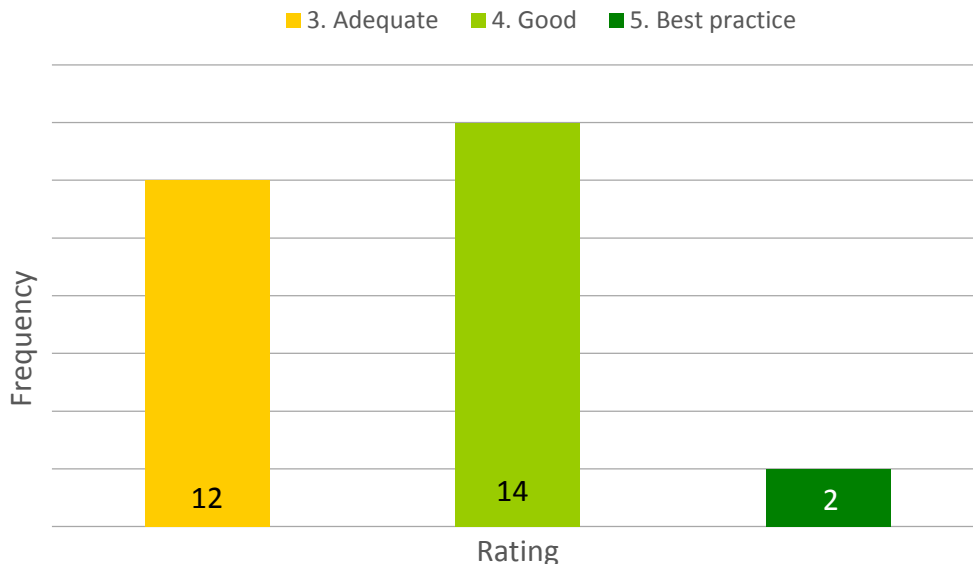
Philippines APPR

The Philippines APPR contains an in-depth, concise discussion of development in the Philippines that provides insight into the challenges over the last 12 months. The context section provides clarity on the level of support for reform by government, and the challenges that could undermine continued growth. It goes on to describe how and why Australia engages in five priority areas and the development challenges it seeks to address. The logic of how the program contributes towards achieving the objectives is clearly reflected in the discussion of progress against each objective.

2.2 Progress assessment

CORNERSTONE QUESTION 2	Does the APPR narrative explain and justify, in a proportionate way, the progress ratings against each program objective?
Subquestion 2a	Are the objectives sufficiently framed at the outset or in the narrative as statements of intent that are measurable (quantitatively and/or qualitatively)?
Subquestion 2b	Does the narrative explain and justify the progress ratings against each objective?
Subquestion 2c	Where appropriate, is the policy dialogue engagement with partners clearly explained?
Subquestion 2d	Is there an appropriately open, balanced discussion of the program's successes, failures and challenges?
Subquestion 2e	Has relevant evidence been drawn from an appropriate range of internal and external sources given the program context?
Subquestion 2f	Are the results appropriately contextualised and the contribution of DFAT made clear?
Subquestion 2g	Are the sources of evidence referenced appropriately?

Figure 6 Frequency of quality ratings against 'progress assessment'



Program areas rate progress towards achieving objectives in the APPRs, using a green, amber, red ‘traffic light’ system.* This requires a frank and balanced assessment of performance, clear objectives, and a clear justification for ratings, including a discussion around the contribution of policy dialogue. Framing a measurable objective involves presenting the objective in the narrative, the PAF or the objective statements themselves in a way that makes clear what is trying to be achieved, and the basis on which performance is judged. For the ratings to be convincing, APPRs need to draw on concrete and contextualised evidence, usually beyond the level of individual investments.

All APPRs are at least satisfactory for this cornerstone question (Figure 6). Most APPRs provide a frank and balanced assessment of performance, with only a few painting an overly rosy picture. Eight of the 28 APPRs however, had one or more objective that was too broad or vague, and did not explain or justify how performance ratings had been derived. Without measurable objectives, it is much more difficult to explain and justify progress ratings. For example, simply stating the sector ‘infrastructure’, as done in one APPR, is inadequate as it is not clear what achievement is intended. Four APPRs were considered best practice in terms of framing objectives (see Box 2 for one example).

Better APPRs explain the basis for the ratings, often in an opening statement in the performance discussion under each objective. This usually involves an analysis of the performance of multiple investments that lead to change at the program level, rather than being based solely on one investment. In some cases, however, singular investments do lead to systemic change (particularly sector wide approaches or large partnerships with the World Bank), and may be an exception. The ratings in the better APPRs are supported by concrete, well-referenced evidence that is beyond investment-level evidence.

Box 2 Good practice in framing objectives

Pakistan APPR: clear objectives and measures of performance

The first objective in the APPR for Pakistan includes a number of specific outcomes. The PAF defines indicators of progress with targets for each of the coming three years. The following example is for one outcome:

Objective 1: Generating economic growth and employment through increased trade and improvements in agriculture and industry

Outcome 1: Poor women and men have increased income through business and employment opportunities and increased value of agricultural produce

Indicators:

- Number of poor farmers and small business owners who have increased their income
- Number of small businesses (% women’s small and medium enterprises) established, restored to operations or expanded to the poor
- Number of jobs (full-time equivalent) created for poor women and men.

Better APPRs – Indonesia for example – are also able to place evidence and results in context so the reader understands their significance. A few APPRs do less well in this regard. For example, an isolated result reported in one APPR, which says, “*It provided community based education classes to*

* Whereby Green = “Progress is as expected for this point in time and it is likely that the objective will be achieved. Standard program management practices are sufficient”; Amber = “Progress is somewhat less than expected for this point in time and restorative action will be necessary if the objective is to be achieved. Close performance monitoring is recommended”; and Red = “Progress is significantly less than expected for this point in time and the objective is not likely to be met given available resources and priorities. Recasting the objective may be required.”

6,715 children, including 960 girls”, provides no indication of the significance of this achievement in the particular country context.

Use of performance assessment frameworks

Almost all lower quality APPRs in terms of progress assessment do not have PAFs from which they draw assessments. A PAF is a document that sets out outcomes, indicators and baselines for the purpose of tracking progress towards program objectives. Good PAFs are underpinned by a clear program logic that explains how aid interventions will achieve objectives and the assumptions underlying this logic.

Programs with better quality PAFs were generally better able to justify progress ratings against objectives. The weaker justifications (and PAFs) fail to explain the basis for the ratings or the connection between what the programs are doing and their claimed achievement against a high-level objective. Other poorer explanations rely solely on investment-level data even when those investments are relatively low level compared to the ambition of the objectives.

Prior to the introduction of the aid program’s new performance framework all programs were expected to have a PAF in place, or a plan for its development, by the end of 2013. As described in the previous APPR quality review, for one-third of programs this did not occur. Under DFAT’s new Aid Programming Guide (released July 2014) it is mandatory that *major* country and regional programs (i.e. those with an estimated total Official Development Assistance value of over \$50 million per annum, or those judged to be high-risk or sensitive) include a PAF as part of their Aid Investment Plans. A PAF is only *recommended* for other country and regional programs.

Both this and last year’s quality reviews of APPRs demonstrate that the credibility of APPRs’ assessment of program performance is enhanced when programs can draw on good quality PAFs. ODE therefore considers that PAFs should be required for *all* programs that produce APPRs.

Contribution of policy dialogue to program progress

Better discussions about policy dialogue make it clear what the dialogue aims to achieve and are more precise about the means of that dialogue. Bangladesh, Laos and Pakistan were all excellent reports in this respect (see Laos example in Box 3).

Box 3 Good practice policy dialogue reporting

Laos APPR

Australia continued in its role as co-chair of the Education Sector Working Group (ESWG), which enabled substantive policy dialogue, particularly on education disadvantage, inclusive education, and transparency in budgeting and planning. Four focal groups were established under the ESWG to enable technical discussions between the Lao Government and development partners (including civil society) to support evidence-based policy-making. Provincial ESWG coordination meetings were planned, but did not occur.

Shaping dialogue through the ESWG were recommendations from the 2013 Mid Term Review of the Education Sector Development Plan (ESDP), to which Australia contributed technical assistance. The review highlighted some of the most pressing issues faced by the sector, such as the government decision to increase civil servant wages (including teacher wages) in 2012–13 and again in 2013–14, putting serious pressure on the non-wage recurrent budget. Inefficiencies related to teacher management and deployment were also highlighted, as teacher salaries consumed 88 per cent of the education recurrent budget for 2013–14, and 94 per cent of the sub-national recurrent budget. The review prompted the Ministry of Education and Sports to take some actions to address this issue, however further substantive action is required.

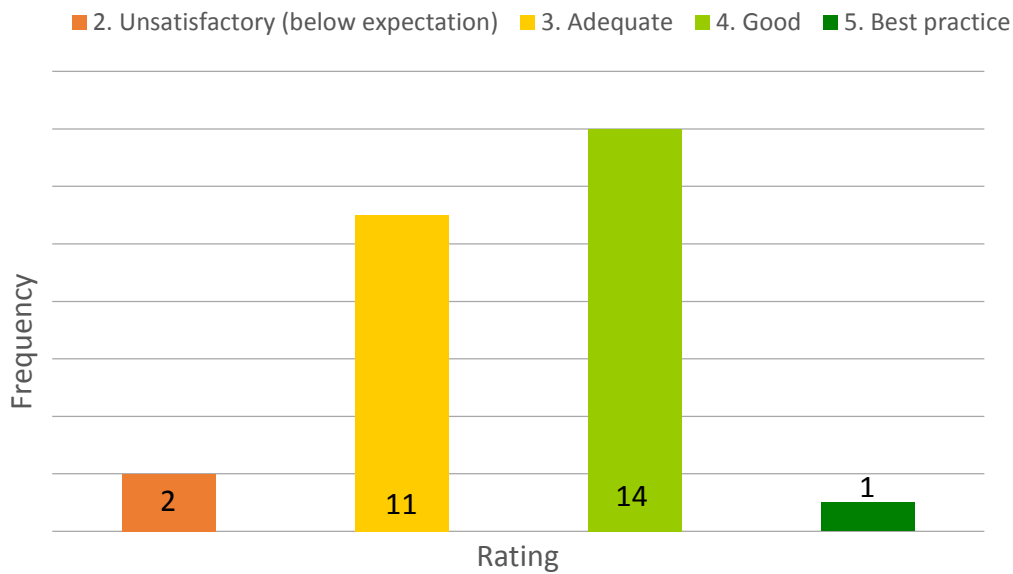
The Government of Laos has requested Australia's help in preparing the next ESDP (2016–2020), an indication of the high regard for Australia's leadership role in the sector. In April 2014, Australia funded, jointly with UNICEF, a workshop as a first stage in developing the new plan. Australia was also asked to help establish a methodology to review the Education Law (2007). Critical issues highlighted for further consultation both within government and with development partners include language of instruction and the length of compulsory education. Australia will pursue this further through the ESWG.

Laos APPR 2013–14

2.3 Partner contribution

CORNERSTONE QUESTION 3	Does the APPR provide a proportionate, balanced discussion of the contribution of partner governments and delivery partners?
Subquestion 3a	For partner governments?
Subquestion 3b	For whole-of-government partners?
Subquestion 3c	For international partners?
Subquestion 3d	For civil society and private sector organisations?

Figure 7 Frequency of quality ratings against 'partner contribution'



APPRs report on the contribution of partner governments and delivery partners towards delivering the aid program objectives. Most reports are at least satisfactory in assessing partner contribution, but two were not (Figure 7).

Whole-of-government partners are mentioned in most APPRs when relevant in describing the program, but their performance is not usually assessed. In regard to assessment of international (multilateral) partners and non-government organisation (NGO) partners, some APPRs make a detailed assessment, particularly of multilateral organisations (as per the guidance), whereas others only mention them as partners. Given the aid program has developed a new partner performance system for key delivery partners (see Target 7 in *Making Performance Count*), expectations for performance reporting in this area would benefit from clarification for future APPR reporting.

Mutual obligations

A section on mutual obligations was introduced to APPRs for the first time in the 2014 reports. This was in response to *Making Performance Count*, which requires mutual obligations to reflect partners' own reform commitments of particular importance to their development as well as Australia's aid investments (see Box 4). DFAT's intent is that APPRs will reflect on **partner government** performance against mutual obligations while the obligations of other partners will be reported elsewhere.

Box 4 Mutual obligations

High-level Target 6

Delivering on commitments: From July 2015, progress against mutual obligations agreed between Australia and its key partner governments and organisations will form part of program performance assessments.

Aid by itself cannot deliver development. We need to move away from traditional donor–recipient relationships to more mature partnerships. Greater mutual accountability is critical in making this transition. While Australia will be held to account for the performance of its investments, partner governments and organisations also need to contribute to shared development goals. Progress by both Australia and its partners in meeting mutual obligations will be assessed and reflected in future budget allocations. The mutual obligations of partner governments and organisations will generally reflect their own reform commitments of particular importance to their development and Australia's aid investments.

Making Performance Count, p.10.

Partner government performance is discussed throughout reports and within the new report section titled 'mutual obligations'. Better reports (such as Burma, Laos, Palestinian Territories, Philippines and Solomon Islands) carefully explain the strengths and weaknesses of partner government systems and how the aid program works in those contexts. Other reports assume knowledge or do not fully explain the relationship between the aid program and partner governments.

Staff preparing APPRs have interpreted the guidance on how to address mutual obligations in a variety of ways. The section is missing entirely in three APPRs. In most Pacific APPRs, reference is made to the mutual obligations in the Pacific Partnerships that are linked to performance payments. The Pacific Regional APPR's mutual obligations draw on core funding arrangements with regional organisations and specific obligations that need to be met. Several APPRs* reflect a common understanding of high-level mutual obligations and specific mechanisms through which they are reviewed. Others discuss mutual obligations at the investment level.†

Two good practice sections on mutual obligations are from the Philippines and Solomon Islands APPRs (see Box 5). These are considered good practice because they contain clear partner government obligations and are able to describe progress against them, drawing on sound evidence.

* Afghanistan, Bangladesh, Burma, Laos, Nepal, Palestinian Territories, Philippines, Timor-Leste and Vietnam.

† Mekong Water Resources Program, Mongolia, South Asia Regional, Sri Lanka and Sub-Saharan Africa.

Box 5 Good practice reporting of mutual obligations

Philippines APPR

The Philippines APPR explains how the Philippine Government articulates its development agenda in the Philippines Development Plan 2011–2016, which forms the anchor for the bilateral development agreement between Philippines and Australia, through the Statement of Commitment signed in March 2012. The Philippine Government is making significant progress in its eight commitments under the Statement of Commitment; the APPR lists the eight commitments and gives a detailed progress report against each.

Solomon Islands APPR

The Solomon Islands APPR carefully explains how Solomon Islands Government–donor engagement takes place, through the Core Economic Working Group (CEWG). It describes major, concrete achievements; for example, the Solomon Islands Government has approved the Public Finance Management Reform Roadmap, published a Budget Outcome Statement and enacted the Constituency Development Fund Act, with regulations under development. It explains that Australia and Solomon Islands have formally agreed mutual outcomes in the health, education, transport and governance sectors and reports against these in detail. Independent assessment of 2013 performance confirms that, overall, insufficient targets were reached, making a performance-linked aid payment in 2013 unlikely. The APPR details the arrangements and performance under the CEWG and the link to incentive payments available under that mechanism.

The varied interpretation across APPRs on how to report on mutual obligations reflects the newness of this requirement and the associated lack of clarity in the APPR guidance (the template and good practice note). This was the first instance of DFAT’s dealing with mutual obligations and will take some time to bed down. In future, mutual obligations will be articulated in Aid Investment Plans and reported against in subsequent APPRs—this should add clarity to what is required in APPRs. In addition, the next iteration of DFAT’s APPR guidance should further clarify mutual obligation reporting expectations, including direction on what to include/exclude, how to assess positive versus negative performance, and whether just partner performance or *mutual* performance against obligations is to be assessed.

2.4 Management consequences

CORNERSTONE QUESTION 4

Are the management consequences clearly linked to the ratings narratives and the identified risks to future progress?

Subquestion 4a

Is there a logical connection between the management consequences section and the risks to future progress flagged in the other sections of the APPR?

Subquestion 4b

Are the specific management consequences framed so they are relevant to program strategy and management and are also actionable within the reporting period? Are they also prioritised?

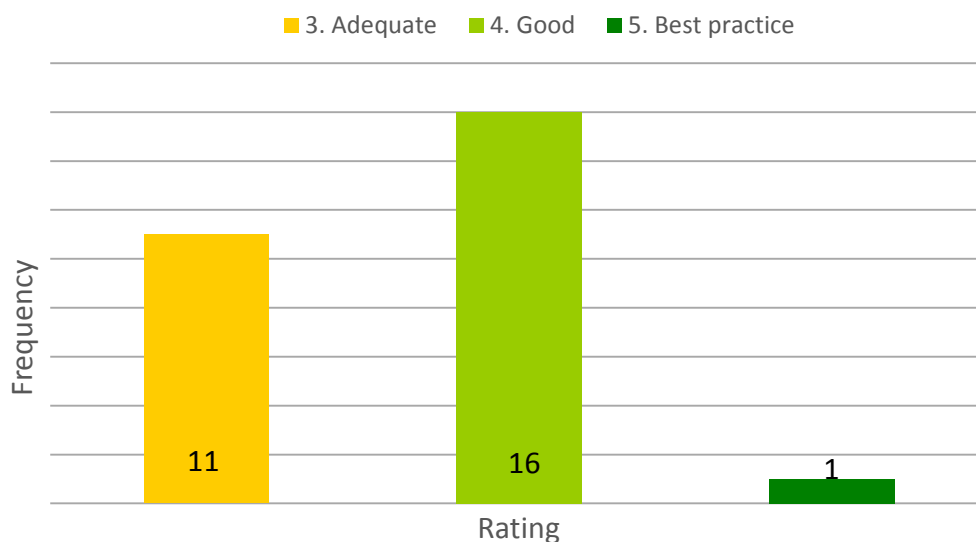
Subquestion 4c

Are lessons from the experience of implementing the previous year’s management consequences noted and acted upon in this year’s consequences section?

Previous APPR quality reviews reinforced the importance of program teams using the APPRs as management tools, and not just for accountability purposes. It is clear from this year’s (Figure 8) and last year’s quality assessment that progress towards this goal is well on track. Almost all APPRs provide considered management consequences, linking these to the prior year’s responses and quality ratings in this year’s APPRs. The main point of difference between good practice and adequate practice is in the specific and actionable nature of the management consequences, as well as being pitched at an appropriately strategic level (as opposed to listing each small action to be taken at investment level).

The quality of this year’s responses to management consequences may be slightly lower than last year’s. However, it is perhaps unsurprising that with the budget uncertainly prevailing at the time they were authored, some APPRs are less concrete about management actions than they have been in previous years. This is because a large proportion of management responses have budget implications. Overall, they still indicate a reasonably actionable set of management consequences usually pitched at an appropriate level.

Figure 8 Frequency of quality ratings against ‘management consequences’



2.5 Other observations

Program size

APPRs for smaller programs—less than \$50 million—tend to be lower quality (Appendix C). This is consistent with the finding in the 2013 Quality Review of APPRs that smaller programs may produce lower quality APPRs. This is likely to be the result of smaller programs having fewer staff. Under DFAT’s management system, often a single member of staff, more junior than those allocated to larger programs, covers multiple sectors and manages the program as well as reporting on it. If this is the case then the Kiribati team, particularly, should be congratulated for producing a high-quality APPR.

Gender equality and empowering women and girls

A highlight for several APPRs this year was a strong and strategic focus on gender equality and empowering women and girls. This was particularly true for the Burma, Fiji and PNG APPRs, though gender equality issues were well discussed in most APPRs.

Regional APPRs

Regional programs in the Australian aid program have long suffered from a lack of clarity in how they intersect with bilateral programs. This year’s APPRs do a better job in separating truly regional work from multi-country programs. The Pacific Regional APPR contains a series of annexes relating to the bilateral achievements from the regional program. In other APPRs—Nepal for example—the bilateral assessment includes regional (and global) contributions. Bilateral contributions from regional programs should be consistently reported in the bilateral APPRs as they form part of Australia’s aid contribution to achieving bilateral objectives.

3 Performance benchmarks

Making Performance Count establishes the requirement that country and regional programs set performance benchmarks to assess progress towards objectives, and to inform annual budget allocations. Aid Investment Plans, due September 2015, will include longer-term performance benchmarks, covering the duration of the plans (generally four years in line with budget forward estimates). In the meantime, country and regional programs identified a small set of temporary benchmarks in 2013–14 APPRs to assess progress in 2014–15. This ‘pilot’ of performance benchmarks allows learning to inform the longer-term benchmarks to be included in Aid Investment Plans. The following analysis contributes to this learning.

The 2013–14 APPR guidance suggests an indicative number of benchmarks for different sized programs*, and states that benchmarks should:

- › be relevant indicators of progress, providing a clear link between country program aid objectives and benchmarks
- › be predominantly within Australia’s control to achieve
- › be achievable and measurable within a 12-month period (however, programs could identify longer term benchmarks that could be carried over to Aid Investment Plans)
- › draw on existing monitoring and data collection efforts
- › reflect program size and sectoral breadth, focusing on the most significant financial and strategic investments, including (at programs’ discretion) the activities of other government departments where they deliver a significant amount of Australian Official Development Assistance.

This chapter examines the characteristics and quality of the temporary benchmarks.

3.1 Performance benchmark characteristics

The 28 APPRs from 2013-14 contain 218 performance benchmarks of seven different types (Table 2). The majority of benchmarks were classified as one type, but some were given more than one classification. For example, we classified the benchmark “40 partnerships with private sector in place” as both an ‘output’ and ‘strengthen partnerships’ type of benchmark. The most frequently used benchmarks were outputs, followed by outcomes and aggregate results (Figure 9). Programs generally used a mix of the different types.

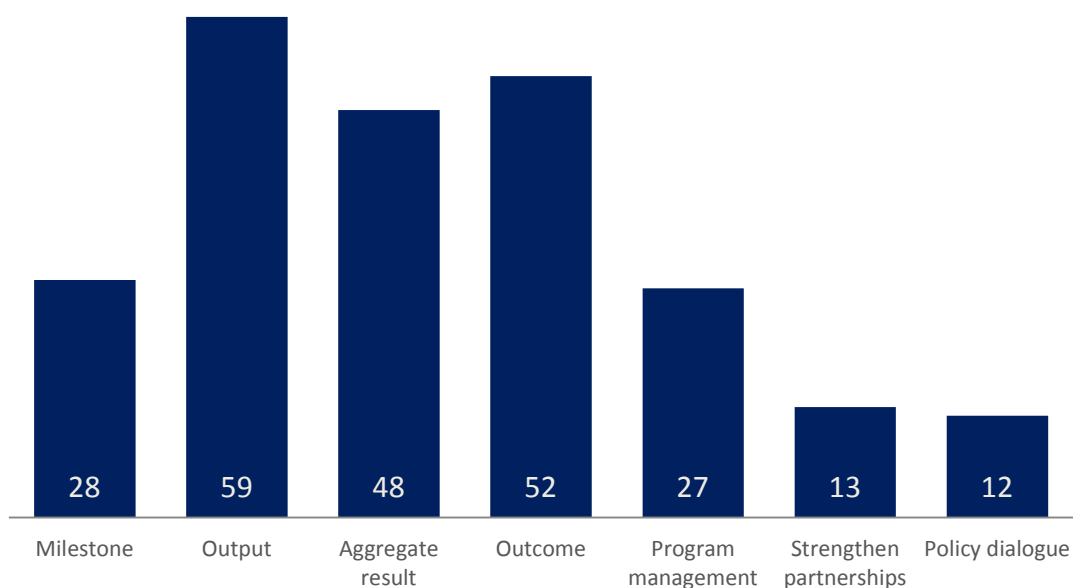
* The guidance suggests that programs with a total value of less than \$50m in 2013–14 identify three to six performance benchmarks; \$50m–\$300 programs identify six to eight benchmarks; and program valued at more than \$300m identify up to ten benchmarks.

Table 2 **Types of performance benchmarks in APPRs**

Benchmark type	Definition	Example
<i>Milestone</i>	A significant point in a contract/program	Commencement of civil works on National Highway One
<i>Output</i>	Good/service delivered by the intervention	Eight irrigation infrastructure improvements completed
<i>Aggregate result*</i>	Intended results of the program, often relating to beneficiaries	10,000 children receive measles immunisation
<i>Outcome</i>	Changes that the interventions aim to deliver in the communities they are working in Short term: changes in knowledge, skills, attitudes and motivation Medium term: behaviour change	80% of trained teachers more confident in their ability to teach new curriculum 10% increase in men demonstrating health seeking behaviours
<i>Program management</i>	Related to administration of the aid program; may cover aspects of efficiency, or how the program is targeted/aligned	Increase the percentage of program budget invested in private sector development, aid for trade and infrastructure from 40 per cent to 60 per cent
<i>Strengthen partnerships</i>	Focus on the relationship between DFAT and partners (partner government, NGOs, multilaterals)	Host annual development cooperation partnership talks
<i>Policy dialogue</i>	Australia discussing policy with partner governments	Agree with partner government on Australian/other donor assistance to support a program of reforms to improve public financial management

* Note that this benchmark type is unrelated to the DFAT Aggregate Development Results, which are a set of 20 specific measures that each monitored aid investment is required to report against on an annual basis, and which contributes to the set of department-wide results from the aid program that are reported against on an annual basis.

Figure 9 The most common types of performance benchmarks identified in 2013–14 APPRs were outputs, outcomes and aggregate results



Quality of performance benchmarks

APPRs did not provide enough information to assess the achievability of the performance benchmarks or to measure the extent of stretch or ambition of the targets. The absence of baselines was the biggest hindrance in this regard. In spite of this limitation, each program’s benchmarks were assessed as being higher, adequate or lower quality (Table 3).* The Cambodia, Pakistan and Timor-Leste APPRs had particularly strong sets of benchmarks.

The assessment was based on analysis of:

- › whether DFAT’s contribution to achieving the benchmark was distinguishable (from partner governments) and significant
- › alignment with APPR objectives
- › sectoral coverage, comparing the number of benchmarks and proportion of budget allocated to each sector
- › whether benchmarks reflected an orientation to new aid policy priorities, specifically aid for trade, private sector investment, and women and girls
- › quality, content and specificity of the benchmarks, considering whether they meet the *specific, measurable, achievable, realistic and time-bound* (SMART) criteria.

* The review team developed the initial framework for assessing performance benchmarks based on consideration of DFAT policy documents (e.g. APPR template and good practice note). The framework was then finalised in collaboration with ODE and Aid Management and Contracting Division.

Table 3 Quality of program benchmarks by country or regional program

Quality category	Country or regional program
<i>Higher quality benchmarks</i>	Bangladesh*, Cambodia*, Laos*, Pacific Regional, Pakistan*, Palestinian Territories, Solomon Islands*, Timor-Leste*, Vietnam*
<i>Adequate quality benchmarks</i>	Afghanistan*, ANCP, Burma*, East Asia Regional*, Fiji, Indonesia, Mekong Water Resources Program*, Mongolia*, Nepal, Philippines*, South Asia Regional, Sub-Saharan Africa, Tuvalu, Vanuatu
<i>Lower quality benchmarks</i>	Kiribati, PNG, Samoa, Sri Lanka, Tonga

* Programs with a PAF.

Characteristics of higher and lower quality performance benchmarks

Three main factors affected the quality of performance benchmarks: the use of a PAF; program size and geography; and specificity and mix of benchmark types.

1. Thirteen country/regional programs have a PAF (Table 3). While having a PAF is no guarantee of stronger benchmarks, there does appear to be a relationship between having a PAF and identifying an adequate or better set of benchmarks.
2. Programs with smaller budgets—particularly in the Pacific—had the most trouble producing good quality benchmarks. This likely relates to the resourcing issues mentioned above. Pacific programs constituted four of the five that fell into the category of lower quality benchmarks (Table 3). Three other Pacific programs were found to have adequate quality benchmarks, and one, Solomon Islands, was the standout among Pacific programs, with higher quality benchmarks. On balance, programs in the Pacific may require further support to identify an appropriate set of performance benchmarks in Aid Investment Plans.
3. Higher quality sets of benchmarks benefited from the specificity of measures framed as aggregate results and fewer milestones. In contrast, lower quality benchmarks suffered from a higher proportion of poorly defined outcomes.

Characteristics common to the nine programs with higher quality performance benchmarks and the five programs with lower quality performance benchmarks are listed in Table 4.

Table 4 Characteristics of good and poor quality performance benchmarks

Quality category	Characteristics
<i>Higher quality benchmarks</i>	<ul style="list-style-type: none">> Align to objectives and strategic priorities outlined in the APPR.> Provide meaningful measures of the aid program; that is, DFAT's contribution is distinguishable (from partner governments) and significant (in relation to progress towards achieving objectives).> Are commonly drawn from a good PAF that sets out a trajectory of change in relation to each of the APPR objectives and identifies DFAT's contribution to these objectives.> Have a balance of types of benchmarks, with a bias towards aggregate results.> Are clear, specific and measurable.> Provide a baseline and/or targets.
<i>Lower quality benchmarks</i>	<ul style="list-style-type: none">> Include too many benchmarks.> Are not specific or assessable.> Combine multiple concepts or measures.> Lack alignment with APPR objectives.> Lack balance with sectoral priorities.> Have a high proportion of outcome measures that are not specific.> Are drawn from investment-level documents (where these are already reported in Aid Quality Checks they may not be ideal as benchmarks).

3.2 Performance benchmarks as measures of performance

DFAT intends to use the progress against performance benchmarks to inform annual budget allocations, so it is important that they are robust performance indicators. Although the majority of performance benchmarks in 2013–14 APPRs are of adequate or better quality against the assessed criteria, this does not mean that on their own they serve as good measure of program performance.

ODE consider that performance benchmarks can play a role in assessing program performance; but that the benchmarks alone will not enable the most credible assessment, particularly of *relative* program performance. Rather, it would be better for the department to consider performance against the benchmarks in combination with other measures, especially APPR 'traffic light' ratings, to inform assessments of program performance.

4 How well do APPRs reflect changes in aid program performance?

APPRs rate program progress towards objectives with a ‘traffic light’ system (Figure 10).

Figure 10 APPR ‘traffic light’ system

GREEN	Progress is as expected for this point in time and it is likely that the objective will be achieved. Standard program management practices are sufficient.
AMBER	Progress is somewhat less than expected for this point in time and restorative action will be necessary if the objective is to be achieved. Close performance monitoring is recommended.
RED	Progress is significantly less than expected for this point in time and the objective is not likely to be met given available resources and priorities. Recasting the objective may be required.

This distribution of APPR ratings over the past three years is shown in Figure 11.

Figure 11 Distribution of APPR ratings against objectives 2011–12 to 2013–14

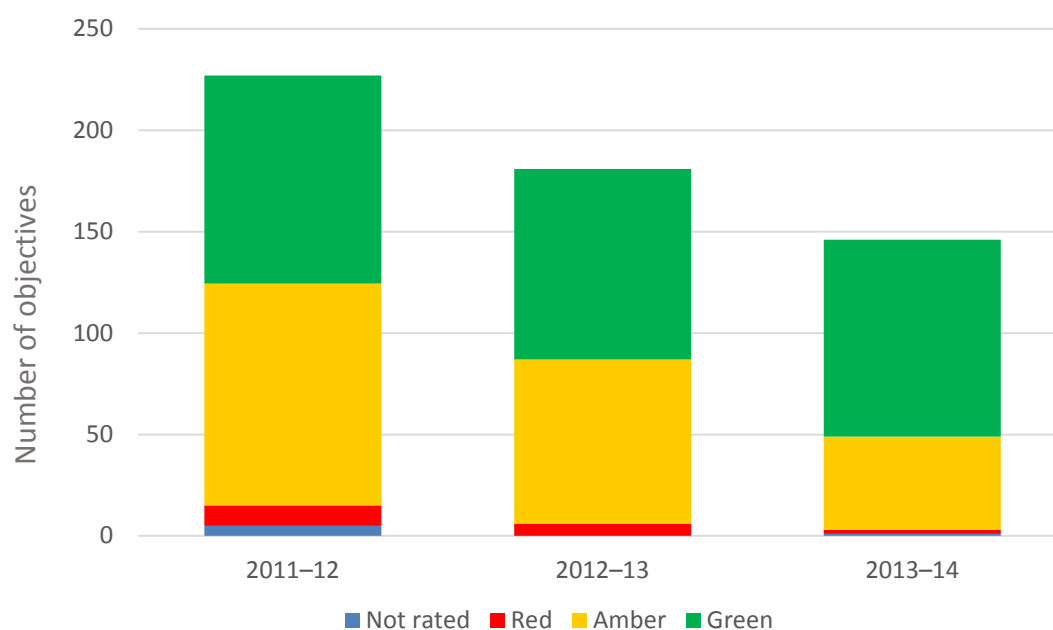


Figure 11 shows a large decline in the number of objectives in APPRs since 2011–12 (from 227 to 146). The number of objectives rated amber has almost halved, and objectives rated red—though always low—have almost disappeared. The number of objectives rated green have changed only slightly. If taken at face value, the greater relative prominence of green ratings in 2013–14 APPRs (green representing 68 per cent of total ratings, compared with 52 per cent 2012–13 APPRs), combined with the reduced prominence of amber and red ratings, could be read as indicating that the aid program is now much better delivering on its objectives than it did two years ago. On this, it is relevant to note the following point ODE made in our quality assurance of the *Performance of Australian Aid 2013–14* report:

*We caution against putting significant weight on year-on-year comparisons of APPR ratings, especially given the long timeframes usually associated with achieving substantial development results. **

To better understand the reasons for this trend ODE focused on changes between the 2012–13 and 2013–14 APPRs.

The first change we looked at was why there were 36 fewer objectives in 2013–14. Firstly, policy changes meant fewer APPRs were completed in 2013–14.[†] The reduction in APPRs accounted for just over half (20 out of 36) of the reduction in objectives, and had an equal effect on amber and green ratings.

Secondly, 26 objectives were dropped, and 10 new objectives added. The net effect of removing or adding objectives was the reduction of 15 amber and red ratings—but only three green ratings. We could not confirm whether this was because investments supporting objectives rated amber or red were more likely to be discontinued than those that were going well.

We then looked at any changes in ratings given for the objectives that remained relatively constant between 2012–13 and 2013–14. This was 134 of 145, or 90 per cent of 2013–14 objectives. Ratings against 82 of the 134 objectives were unchanged, and the remaining 52 either showed improved or deteriorated ratings. Of these, ratings against 15 objectives declined, and 37 improved.

Looking at the 37 cases of reported improvements against program objectives in the latest APPRs, ODE analysis found that:

- APPR narratives provided reasonable supporting evidence to justify 22 of these changes
- in six cases ratings improved because program areas introduced more realistic and specific performance expectations, or altered the scope of the objective
- for the remaining nine cases, improvements in ratings were not well substantiated because the objectives and/or indicators were too vague or the evidence presented in APPRs too weak.

Overall, this analysis provides further substantiation to ODE's preliminary findings, as reported in the *Performance of Australian Aid 2013–14*, where we suggested, that:

To allow better understanding of change in country and regional program performance, it will be important to ensure consistency in indicators and interpretation of progress over time.[‡]

Following from this, changes in APPR objectives, and the way progress against them is measured, must be taken into account if ratings against APPR objectives are to be used to claim changes in program performance over time.

If the department wishes to strengthen the capacity of APPRs to measure performance over time, country and regional programs should ensure stability of objectives, and indicators from year to year (including transparency of how progress against them is measured). Encouraging

* *Performance of Australian Aid 2013–14*, DFAT, Canberra, 2015, p.71.

† In 2012–13, APPRs were prepared for the Regional Assistance Mission to Solomon Islands (RAMSI); North Pacific; Middle East and North Africa; Latin America; and the Caribbean—but not in 2013–14. (Three mini APPRs—for Bhutan, India and Maldives—which were counted as one APPR were also completed in 2012–13, but were not included in the counts of that year's APPR objectives.)

‡ As per citation at * above.

this stability, Aid Management and Contracting Division could monitor changes to objectives and indicators in APPRs each year.

Appendices

Appendix A: Terms of reference (excerpt)

Objectives

1. In quality assuring and analysing the 2013–14 APPRs, ODE's 2014 Quality Review of APPRs will:
 - a) Comment on the coverage and key characteristics of the country and regional program-level performance benchmarks proposed in the 2013–14 APPRs
 - b) Report on the reasonableness of the performance narratives in the 2013–14 APPRs and the use of evidence.

Approach

2. Evidence for the 2014 Quality Review of APPRs will be gathered from the following sources:
 - a) Documents concerning the emerging performance benchmarks and simplified reporting system
 - b) 2013–14 APPRs
 - c) 2012–13 APPRs (to compare performance and quality)
 - d) APPR good practice notes and templates
 - e) Previous ODE APPR quality reviews and working papers.
3. The quality review consists of three main tasks:
 - a) Review the performance benchmarks proposed in all 2013–14 APPRs, focusing on their coverage, quality and character
 - b) Assess the 2013–14 APPRs against the following quality criteria, consulting with ODE on the assessment method:
 - Does the APPR provide a clear, balanced description of program relevance within the country/regional context and the new aid policy framework?
 - Does the APPR narrative explain and justify, in a proportionate way, the progress ratings against each program objective?
 - Does the APPR provide a proportionate, balanced discussion of the contribution of partner governments and delivery partners?
 - Are the management consequences clearly linked to the ratings narratives and the identified risks to future progress?
 - c) Write the 2014 Quality Review of APPRs.

Appendix B: APPR quality review assessment questions

A) CONTEXT AND STRATEGIC ORIENTATION

Cornerstone question: Does the APPR provide a clear, balanced description of program relevance within the country/regional context and the new aid policy framework?

Subquestions:

- a. Does the APPR acknowledge, respond to and reflect the new aid policy?
- b. Does the context analysis explain the key development challenges/needs of the country/region, with an emphasis on changes that have occurred during the reporting period? Is this analysis used to explain the Australian program objectives?
- c. If there is a stated strategy or 'theory of change' does this explain how program objectives will be achieved, and does this explanation reference the country/regional context analysis? In the absence of a stated strategy or theory of change, is some other explanation for achieving program objectives, given the context, provided?
- d. Does the report make clear the proportion of Australia's aid in relation to other donors and national indicators, and the program and strategy implications of that?

B) ASSESSMENT OF PROGRESS TOWARDS OBJECTIVES

Cornerstone question: Does the APPR narrative explain and justify, in a proportionate way, the progress ratings against each program objective?

Subquestions:

- a. Are the objectives sufficiently framed at the outset or in the narrative as statements of intent that are measurable (quantitatively and/or qualitatively)?
- b. Does the narrative explain and justify the progress ratings against each objective?
- c. Where appropriate, is the policy dialogue engagement with partners clearly explained?
- d. Is there an appropriately open, balanced discussion of the program's successes, failures and challenges?
- e. Has relevant evidence been drawn from an appropriate range of internal and external sources given the program context?
- f. Are the results appropriately contextualised and the contribution of DFAT made clear?
- g. Are the sources of evidence referenced appropriately?

C) PARTNER CONTRIBUTION

Cornerstone question: Does the APPR provide a proportionate, balanced discussion of the contribution of partner governments and delivery partners?

Subquestions:

- a. For partner governments?
- b. For whole-of-government partners?
- c. For international partners?
- d. For civil society and private sector organisations?

D) MANAGEMENT CONSEQUENCES

Cornerstone question: Are the management consequences clearly linked to the ratings narratives and the identified risks to future progress?

Subquestions:

- a. Is there a logical connection between the management consequences section and the risks to future progress flagged in the other sections of the APPR?
- b. Are the specific management consequences framed so they are relevant to program strategy and management and are also actionable within the reporting period? Are they also prioritised?
- c. Are lessons from the experience of implementing the previous year's management consequences noted and acted upon in this year's consequences section?

After each APPR was assessed against the above questions, the review team also commented on:

- a. the overall quality of each APPR
- b. other observations (e.g. reflections on compliance with guidance and emerging themes coming from assessments).

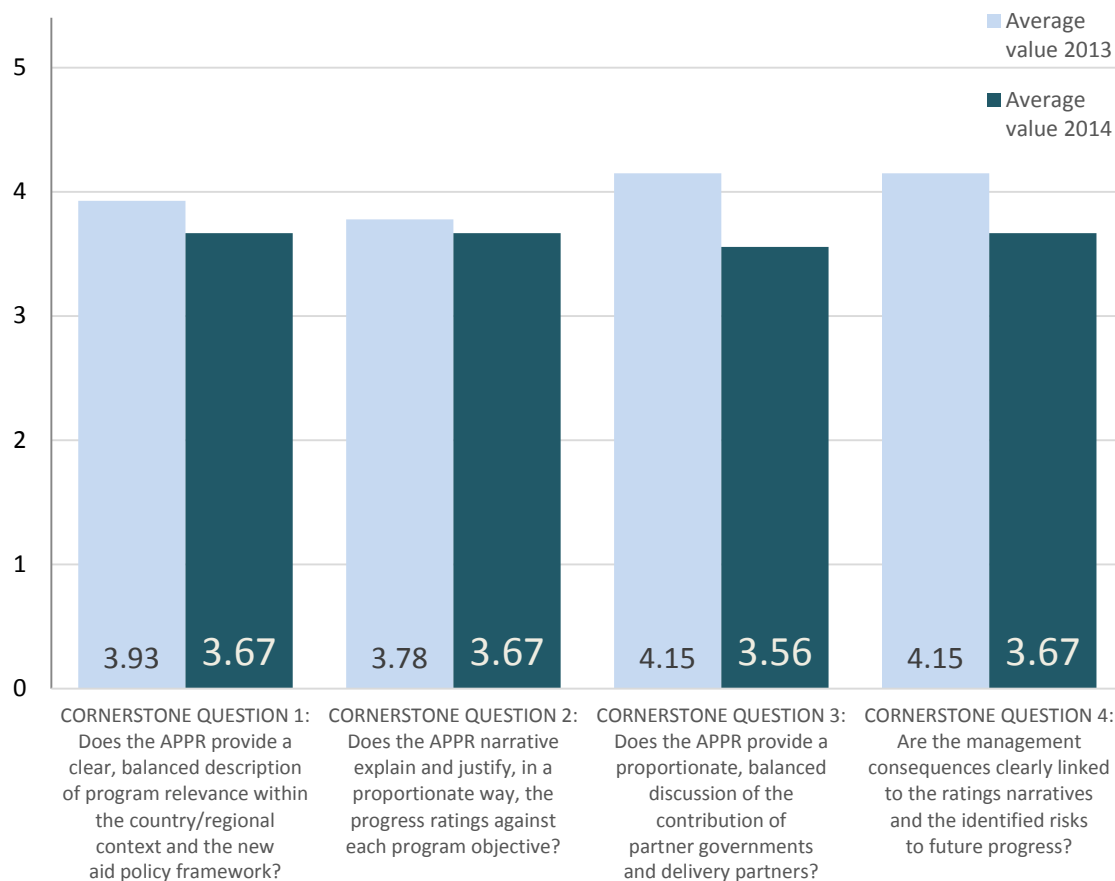
Appendix C: APPR quality by size and whether program has a performance assessment framework

Table C1 APPR quality by program size and PAF*

Quality category	Country program size is highlighted: less than \$50 million, between \$50 million and \$300 million over \$300 million. (*programs with a PAF)	
<i>High quality</i>	Pakistan*	Solomon Islands*
	Bangladesh*	Burma*
	Palestinian Territories	Kiribati
	Philippines*	Timor-Leste*
<i>Good quality</i>	Indonesia	Vanuatu
	Laos*	Afghanistan*
	Mekong Water Resources*	Cambodia*
	South Asia Regional	East Asia Regional*
	Sri Lanka	Pacific Regional
	Vietnam*	PNG
	Samoa	Tonga
	Sub-Saharan Africa	Tuvalu
<i>Adequate quality</i>	ANCP	Mongolia*
	Fiji	Nepal

Appendix D: Cornerstone question assessment comparison over two years

Figure D1 Comparison of ratings on cornerstone questions for 2014 and 2013 APPR Quality Reviews



Appendix E: Summary tables of quality ratings against all cornerstone questions and subquestions

Figure E1 Context and strategic orientation – frequency of all quality ratings

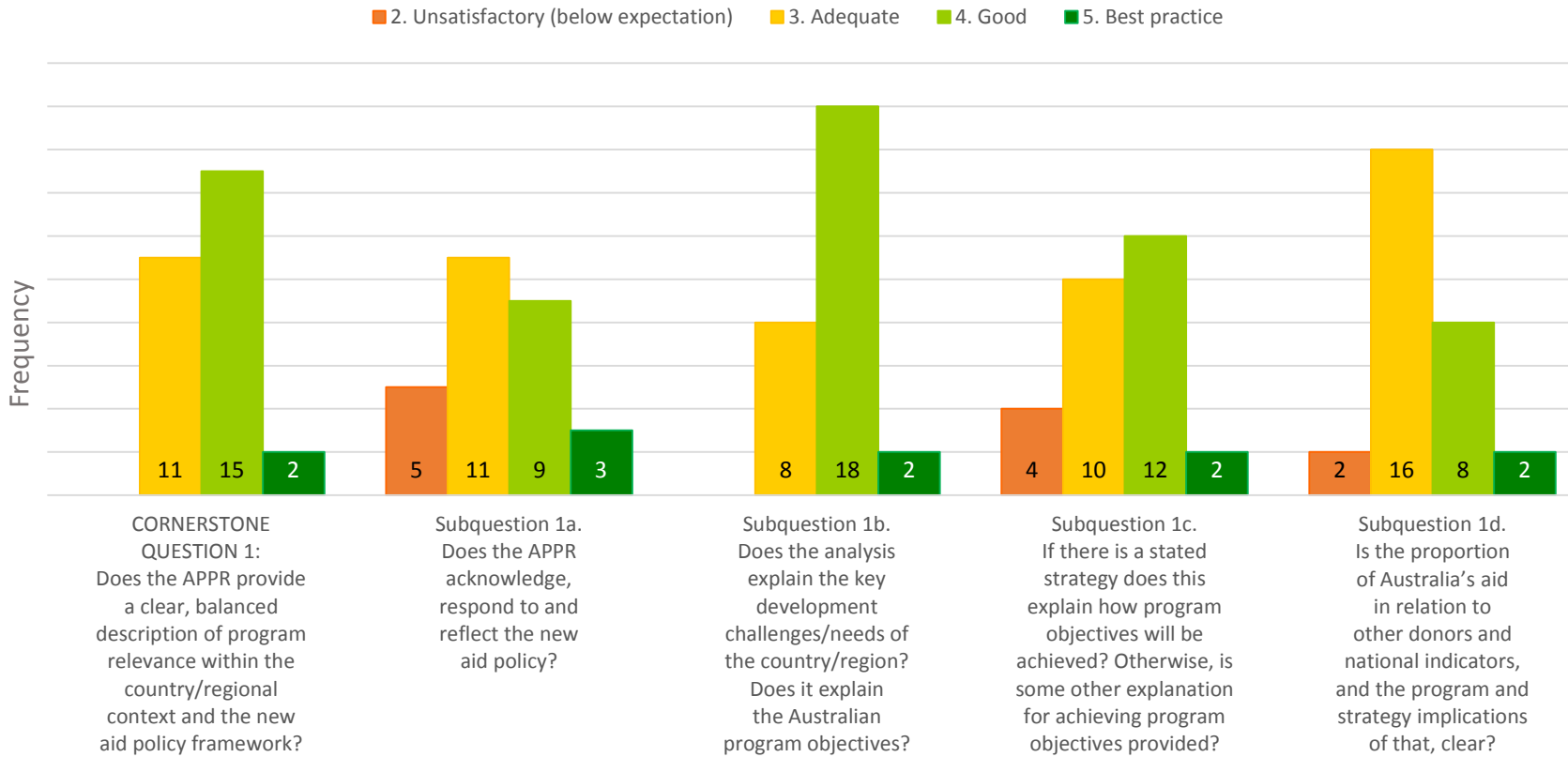


Figure E2 Progress assessment – frequency of all quality ratings

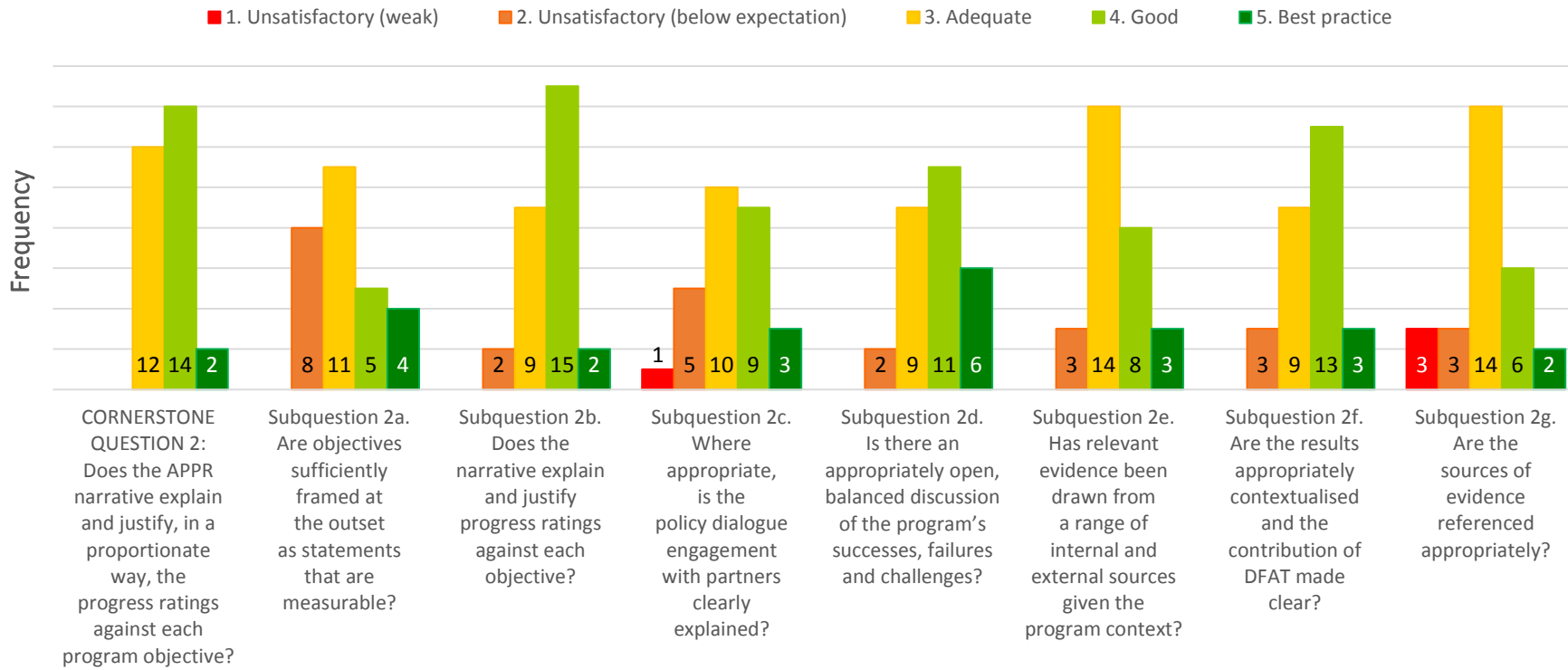


Figure E3 Partner contribution – frequency of all quality ratings

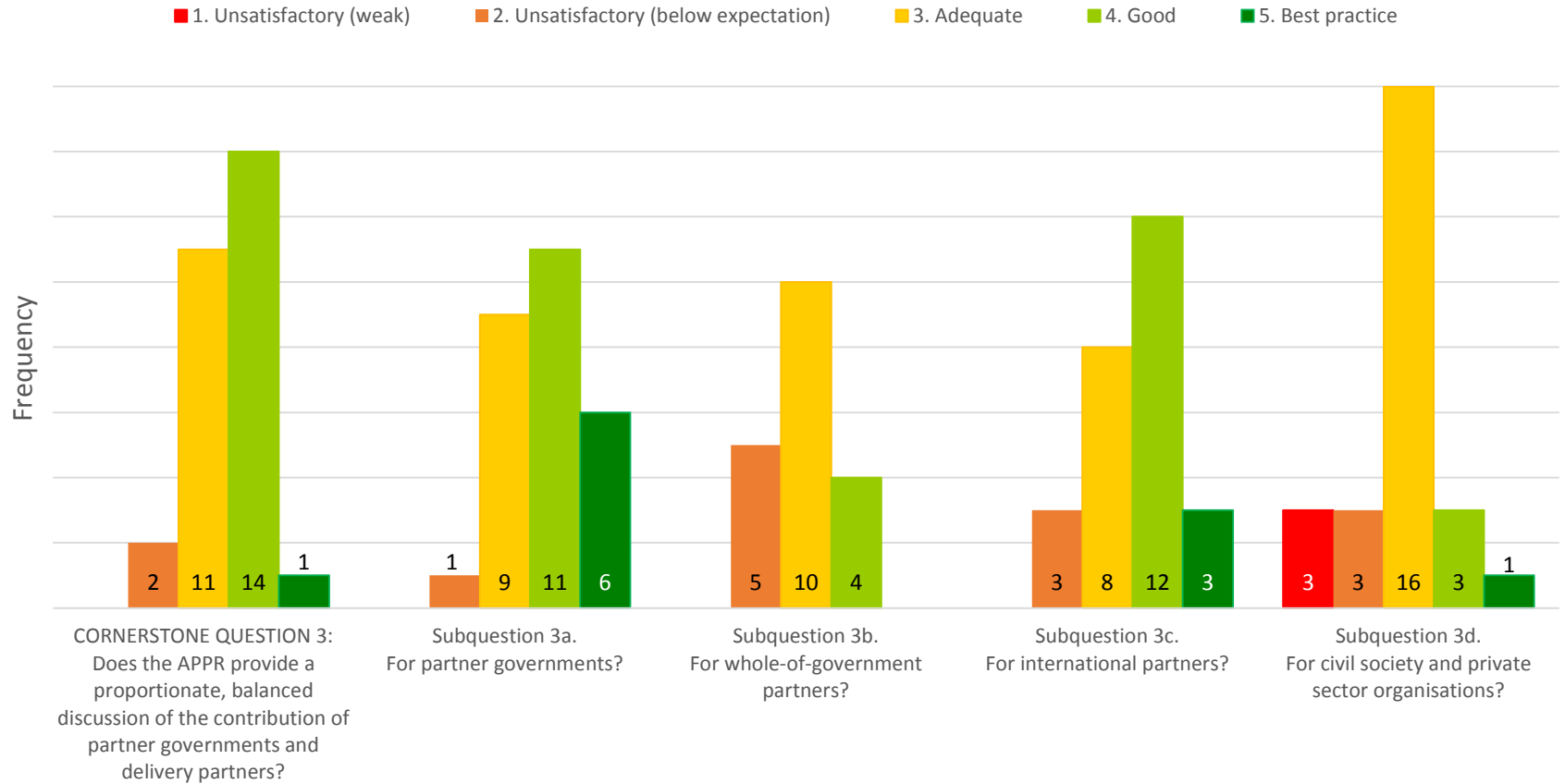
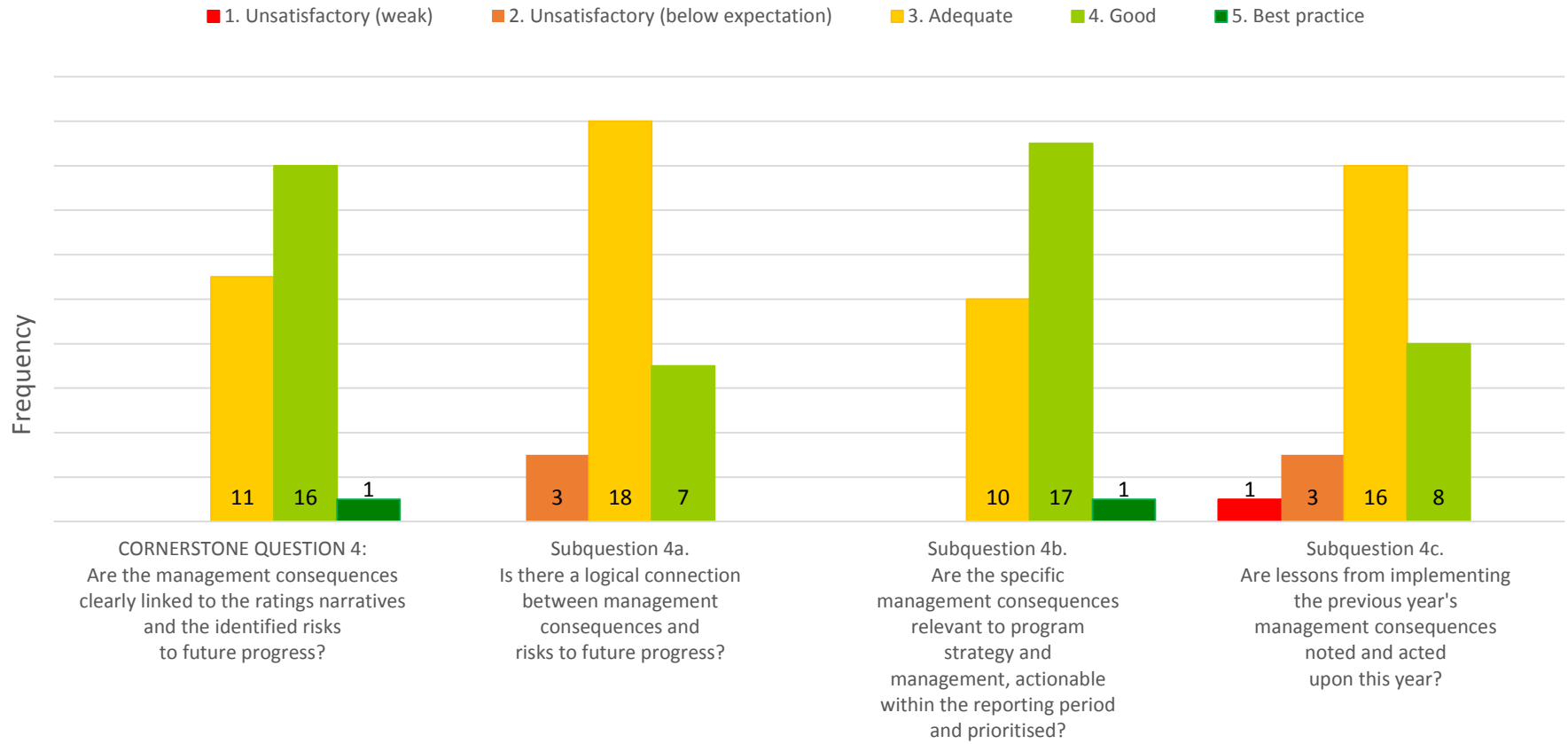


Figure E4 Management consequences – frequency of all quality ratings



Abbreviations

ANCP	Australian-NGO Cooperation Program
APPR	Aid Program Performance Report
CEWG	Core Economic Working Group
DFAT	Department of Foreign Affairs and Trade
ESWG	Education Sector Working Group
NGO	non-government organisation
ODE	Office of Development Effectiveness
PAF	performance assessment framework
PNG	Papua New Guinea