2024 Review of the Quality and Use of DFAT Evaluations

Final Report

Submitted to the
Department of Foreign Affairs and Trade
by Bluebird Consultants

30 JANUARY 2025

bluebird consultants

Contents

KEY FINDINGS **3**

INTRODUCTION **5**

BACKGROUND **5**

PURPOSE **5**

OBJECTIVES **6**

SCOPE **6**

INTENDED USE AND INTENDED USERS **6**

REVIEW QUESTIONS **7**

METHODOLOGY **7**

HOW PROGRAM EVALUATIONS ARE USED **9**

1. WHAT THE LITERATURE TELLS US 9
2. WHAT WE FOUND THROUGH THIS REVIEW 10

KEY ENABLERS AND BARRIERS TO THE USE OF PROGRAM EVALUATIONS **12**

1. WHAT THE LITERATURE TELLS US 12
2. WHAT WE FOUND THROUGH THIS REVIEW 14

Correlation between use and evaluation/ issue prioritisation *14*
Correlation between use and the evaluation process *15*
Correlation between use and evaluation quality *16*
Correlation between use and systems that facilitate use *17*
Correlation between use and organisational evaluation capability *18*
Correlation between use and evaluator *19*

ANNEX 1 – REVIEW PLAN **20**

ANNEX 2 – SAMPLED EVALUATIONS **30**

ANNEX 3 – FULL LIST OF EVALUATIONS **32**

ANNEX 4 - SUMMARY OF ORGANISATIONAL EVALUATION CAPABILITY SURVEY RESULTS **37**

Key findings

DFAT's Development Evaluation Policy identifies use as the driving force behind DFAT evaluations. **This review of the quality and use of DFAT evaluations examines how evaluations were used during the 2023-24 reporting period and to what extent**.

All 46 DFAT evaluations completed between January 2022 and June 2023 were analysed using quantitative data. A sample of 9 evaluations were also analysed using qualitative information. These evaluations covered the range of geographies, sectors, and modalities DFAT is engaged in. Evaluation and development literature on evaluation use and factors which help, or hinder use was also reviewed.

DFAT staff responsible for managing the investments (ie projects/programs) evaluated reported that almost all of the evaluations had been used (45 out of 46, with the remaining evaluation intended for future use). The primary use was to improve programs, capture and report on progress, and inform decision making. Evaluations are also **used in** a range of **other ways**. Teams are using evaluations to understand their programs or aspects of their programs better. Evaluations are also being used to signal intentions and priorities, and they are being used to inform broader learning outside of the team. For example, evaluations are shared with other donors to inform coordinated country-level support.

Five **key factors were explored**[1] to see how they **enabled or constrained the use of evaluations**. Evaluations were more useful when **DFAT was clear on the issues the evaluation should examine**. Additionally, evaluations were valued when they provided a structured way for DFAT to engage with stakeholders around the key issues. **Higher quality evaluations were used more. However, staff had a broad definition of what a quality evaluation is**, informed by their perceptions of the quality of the team, the evaluation process, and whether the evaluation addressed the issue it was commissioned for. Previous quality and use reviews used a narrower, more technical definition of quality, with greater emphasis on evaluative methods, evidence base, and the quality of the written report.

Some DFAT systems are supporting evaluation use. Specifically, developing management responses, publishing the reports on DFAT website and linking design to evaluation are helping staff to consider how to implement recommendations, share findings and inform future programming. Additionally, some staff used technical support facilities to help them commission, undertake and review evaluation outputs. **In contrast, other systems are not supporting evaluation use.** There are gaps in making findings easily accessible and ensuing senior management oversight.

One important factor not originally included in the factors for examination was the **quality of the evaluator or evaluation team**. As noted above, staff considered the evaluation team as part of their understanding of quality. Through interviews we heard of the importance of the evaluator or team and how this was linked to use. Good evaluators were described as understanding the political economy, having strong relevant contextual insight, possessing "expert" skills, and having requisite authority to convene forums with key stakeholders. These forums allowed for discussion of critical factors and often support the strengthening of DFAT's partnerships.

1. The five factors were: 1. How issues or programs were prioritised, 2. the process of evaluation, 3 the quality of evaluations, 4. the systems that exist to enable use of evaluation, 5. organisational capability to manage evaluations.

Six recommendations are provided to support DFAT to better use program evaluations in the future:

1. **The Development Evaluation and Assurance Section (EVS) broaden the definition of evaluation use in DFAT's Development Evaluation Policy and share this definition throughout the agency.** A broadened definition would take account of the range of ways programs can use evaluations in addition to the ways already noted in the policy. The broader definition could note the importance of evaluations for learning, signalling intentions and contributing to the field. The policy could also note the importance of using both the evaluation process and outputs.
2. **The Design and Program Advice Section (DPA) broaden the definition of evaluation quality in the DFAT Design, Monitoring and Evaluation Standards (Standard 10).** The current focus in the evaluation standard is on technical quality of the final report. It would be useful to broaden Standard 10 to incorporate dimensions of evaluation process quality. This could include relevant elements such as "The evaluation process enabled positive engagement with key stakeholders; The evaluation process enabled the DFAT management team to learn; The evaluation process informed relevant DFAT decisions".
3. **EVS to continue to promote "top tips" in procuring and managing an evaluation.** EVS currently shares "top tips" with DFAT staff which emphasise the need to plan an evaluation early, signal early to the market and be flexible with timing. These tips help to ensure teams can find appropriately qualified and experienced evaluators and teams with complementary skills sets. It will be useful for EVS to continue to promote these tips to ensure broad reach of these messages especially given turnover in staffing.
4. **EVS to continue to support staff in aspects of their role as an evaluation manager that they may not have done before and are not familiar with.** Staff noted they were least confident in reviewing evaluation plans and reports in relation to DFAT Standards. In contrast, staff were most confident in developing an evaluation ToR, managing an evaluation and working with evaluation implementing partners to meet DFAT standards. EVS can continue to provide and promote support available through the Evaluation Helpdesk in the areas of evaluation management that staff are least familiar with. EVS can also continue to convene Performance and Quality events on these issues where staff and external specialists can share experience and expertise.
5. **EVS to provide examples to share with staff on evaluation findings communication and dissemination.** Staff noted how useful evaluations can be in programming and policy development and decision-making. However, the evaluations were often not shared across teams or with other key stakeholders in the sector. It may be useful for EVS to share examples of effective communication and well disseminated reports. EVS can also provide advice regarding planning for findings communication and dissemination.
6. **EVS to consider the revised definitions of quality and use in its annual review of evaluations.** EVS undertakes annual reviews of evaluation quality and use. The revised definitions noted above can also be applied to future reviews. While this may reduce the comparability of results over time, the use of revised definitions would still be worthwhile.

Introduction

The Development Evaluation and Assurance Section (EVS) of DFAT commissioned Bluebird Consultants in 2024 to review the quality and use of development evaluations. This review follows four previous reviews of program evaluations in 2012, 2014, 2017 and 2023.

Background

The Australian Government committed to strengthen its approach to evaluation and learning in the 2023 International Development Policy. This review provides data that contributes to one of the indicators of the Performance and Delivery Framework that underpins the policy. DFAT has committed to conducting an annual review of the quality and use of evaluations and publicly report on the findings as a measure of the indicator: *Our development cooperation is informed by monitoring, evaluation and learning*.

DFAT's Development Evaluation Policy ('the policy') introduced in 2016 and updated in 2023 identifies use as the driving force behind DFAT evaluations. The policy notes three areas that help to ensure evaluations are used:

- **Prioritisation:** Evaluations focus on the highest priority issues in DFAT programs.
- **Quality:** Evaluations are credible and robust to increase the likelihood of use.
- **Systems:** DFAT has systems in place to facilitate use, such as management responses.

The current review follows the former reviews of quality and use, each of which had a slightly different focus. The 2014 review examined the quality and credibility of evaluations and factors influencing evaluation quality and use. The 2017 review focussed on the impact of the revised Aid Evaluation Policy (2016) on evaluation practice, quality and use. The 2017 review also provided learnings from the evaluations on policy influence, aid capability and gender equality. In 2023 separate reviews examined the quality and use of evaluations, using the same methodology as the previous reviews. The current review has a primary focus on how evaluations are used, including factors that help and hinder use, including whether the three areas noted in the policy (prioritisation, quality and systems) support better use.

In 2023, EVS developed the Evaluation Improvement Strategy (EIS) to:

- deliver commitments referred to in the new International Development Policy to strengthen DFAT's approach to monitoring, evaluation and learning (MEL)
- implement the evaluation capability component of the development capability uplift budget measure, in line with the Department's 10 Year Capability Framework
- respond to the government's Australian Public Service reform agenda integrating high-quality evaluation into all aspects of program and policy development to support evidence-based decisions.

This review is part of and informed by the broader EIS.

Purpose

The review examines the quality and use of DFAT evaluations (with a particular focus on use) to inform implementation of DFAT's Development Evaluation Policy. The review of evaluations assists DFAT in its broader assessments of the performance of Australia's international development program.

Objectives

The review had four objectives:

1. To better understand how independent program evaluations are being used
2. To examine barriers and enablers to use through better understanding:
    1. How issues or programs were prioritised
    2. The importance of the process of evaluation
    3. The importance of the quality of evaluations
    4. The systems that exist to enable use of evaluation
    5. The importance of the organisational capability to manage evaluations
3. To inform approaches to strengthen evaluation use and quality across the department
4. To contribute to evidence that Australia's development cooperation is informed by monitoring, evaluation and learning.

Scope

The review examines all 46 independent program evaluations completed between 1 January 2022 and 30 June 2023. DFAT shifted to reporting on evaluations by financial year under the *Performance and Delivery Framework*. The adjustment to reporting periods meant that the most recent evaluations that could be examined in this review were those completed prior to June 2023. This was to ensure sufficient time for the use of evaluations to occur after their completion. Although the evaluations completed in calendar year 2022 were covered in the previous review (2023), this review deliberately takes a different and more in-depth look at the use of these evaluations in the reporting period (2023-24).

The review draws on assessments of the quality and use of the cohort of evaluations and sampled nine programs for deeper examination (see Annex 1 for the Review Plan, Annex 2 for the sampled programs and Annex 3 for the list of all evaluations). The nine sampled programs covered the range of geographies, sectors and modalities used in DFAT programs.

Intended use and intended users

The primary intended users of this review are:

- The Development Evaluation and Assurance Section (EVS)
- The Design and Program Advice Section (DPA)
- The Development Effectiveness and Enabling Division (PRD)
- The Development Program Committee (DPC)
- DFAT's Executive

More broadly, the Australian public is also an audience.

The findings from the review will be used by these groups to inform the department's approaches, policies, guidance, training and practices in monitoring, evaluation and learning. Specifically, the findings will be used to support DFAT staff to deliver high quality evaluations that are well used.

Performance and Quality focal points, DFAT monitoring and evaluation advisers and DFAT staff involved in commissioning and managing evaluations will have an interest in the review's findings in assisting them to better commission, manage, and use higher quality evaluations.

The findings of the review will be shared with DFAT staff and other stakeholders in the following ways:

- Key findings will be reported in the *Performance of Australian Development Cooperation (2023-24)* which is published on the DFAT website
- The full report of this review, including executive summary, will be published on the DFAT website.

EVS staff and the Bluebird Consultants will present the review process and findings at appropriate forums within DFAT and externally as opportunities arise.

Review questions

The review answers three evaluation questions.

1. How are program evaluations used?
2. What are the key enablers and barriers to the use of program evaluations? Specifically examining correlation between use and:
    a. evaluation/ issue prioritisation
    b. evaluation process
    c. evaluation quality
    d. systems that facilitate use
    e. organisational evaluation capability
3. What actions are recommended to ensure better use of program evaluations?

Methodology

The review was conducted by a team of three comprising two external consultants and an EVS staff member. The composite team was designed to draw on complementary expertise – evaluation and understanding of DFAT systems and processes.

The team conducted a literature review and drew from a survey on use and assessments of the quality of evaluation reports. The team used the survey to develop a purposeful qualitative sample of 9 cases for deeper analysis through interview and administered an organisational evaluation capability survey to each evaluation manager. The sample included three evaluations with the following levels of acceptance of recommendations noted below.

- Low: Some recommendations not accepted or partially accepted
- Moderate: Most recommendations accepted
- High: All recommendations accepted

The sample also considered:

- geography
- size of programs
- partner-managed vs DFAT-managed
- thematic areas
- types of implementation partner

The review team interviewed 1-3 key stakeholders (Program Manager, Evaluation Manager, Senior Manager) from each of the selected evaluations on:

- importance of prioritisation of key issues or programs
- evaluation process
- quality of the evaluation report
- systems that facilitate use

The team analysed the data to identify key findings and provide examples of different types and levels of use. The team used the Qualitative Comparative Analysis (QCA) method to score the presence and absence of predicted factors, whilst also examining the qualitative detail of each case of use.[2]

Data collection and analysis methods used to answer each evaluation question are noted below.

1. **How are program evaluations used?**

Data collection methods

- Review of evaluation literature on factors that enable and constrain evaluation use
- Survey of all program managers of evaluations completed 1 January 2022 – 30 June 2023. Draw from existing data from 2023 review.
- Interviews with stakeholders from a sample of 9 evaluations based on levels of acceptance of recommendations: high, moderate and low. The interviews examined how evaluations have been used as well as what enabled and constrained use. The interviews explored all types of use including by DFAT, implementing partner, government counterparts and others.

Data analysis methods

- Analyse survey responses and categorise into high, moderate and low levels of acceptance of recommendations
- Thematic coding of interview transcripts to identify the main ways evaluations are used
- Identification of main ways evaluations are used

2. **What are the key enablers and constraints to the use of program evaluations?**

Data collection methods

- Review of evaluation literature on factors that enable and constrain evaluation use
- Survey of the evaluation manager of the 9 sampled evaluations on the evaluation capability of their post/unit
- Interview of evaluation manager and other staff involved for the sampled evaluations examining evaluation prioritisation, the evaluation process, perceptions of evaluation quality and systems that facilitate use
- Basic characteristics of evaluations not covered in the 2023 review (those completed January-June 2023), and the investments they relate to, collected from the EVS management system and recorded in assessment template
- Evaluations not covered in the 2023 review (those completed January-June 2023) rated against quality criteria in assessment template by the review team

Data analysis methods

- Thematic coding of interview transcripts to identify key enablers and constraints to use
- QCA to identify the presence and absence of predicted factors and examination of the details of the 9 cases. Analysis of correlation between levels of use and:
  - evaluation/ issue prioritisation
  - evaluation process
  - quality of evaluation
  - systems that facilitate use
  - organisational evaluation capability

[2] For further information see: Simister, N. and Scholz, V., Qualitative Comparative Analysis, INTRAC for Civil Society, https://www.intrac.org/wpcms/wp-content/uploads/2017/01/Qualitative-comparative-analysis.pdf

3. **What actions are recommended to ensure better use of program evaluations?**

Data collection methods

- Data on evaluation use, barriers and enablers collected under Q1-2 above

Data analysis methods

- Testing of preliminary recommendations amongst team and select program managers
- Finalisation of recommendations after EVS review of draft report

How program evaluations are used

1. What the literature tells us

The evaluation literature provides useful insight into how evaluations are used. We outline these findings first before presenting the empirical findings from the review.

Evaluation use was first thought of **instrumentally.** That is, evaluators were concerned that decision-makers use evaluation findings to change what was being evaluated.[3] However, further study showed a wider range of uses of evaluations.[4] Evaluations were used **conceptually** - program staff used findings to understand their program in a new way. Evaluations were used **symbolically** - they were used as a sign of good management or to persuade or influence a partner. And evaluations were used for **enlightenment** - findings added knowledge to the field and were used broadly rather than only on a specific program or policy.

These definitions focused on the use of evaluation findings. However, further study showed that the process of evaluation itself was also responsible for change.[5] People involved in the evaluation learned from the process, both about the program or policy and learning to think evaluatively.[6] This type of use became known as **process use**.

Early definitions of use were focused on individuals. But theorists also began to understand that use could be **organisational as well as individual**.[7] Organisations can also use evaluations in each of the ways listed above.

Evaluations can also be misused. **Evaluation misuse** is when an evaluation is used unethically or inappropriately. Evaluations can be misused during commissioning or during the process of evaluation or findings can be misused.[8] An example of commissioning misuse is when an evaluation is commissioned for political show or as a delaying action. Process misuse includes a user subverting the

[3] Weiss, C. H. (1967, April). Utilization of evaluation: Toward comparative study. In House of Representatives committee on government operations, The use of social research in federal domestic programs, Part III. Washington, D.C.: Government Printing Office, 426-32.

[4] Weiss, C.H. and M.J. Bucuvalas, (1977) "The challenge of social research to decision making," in C. H. Weiss (ed.) Using Social Research in Public Policy Making. Lexington, MA : D. C. Heath.

[5] Patton, M. Q., Grimes, P. S., Guthrie, K. M., Brennan N. J., French, B. D., & Blyth, D. A. (1977). In search of impact: An analysis of the utilization of federal health evaluation research. In C. H. Weiss (Ed.), *Using social research in public policy making*. Lexington, MA: Lexington Books.

[6] Patton, M. Q. (1998). Discovering process use. *Evaluation*, 4(2), 9.

[7] Cousins, J. B., Goh, S. C., Clark, S., & Lee, L. E. (2004). Integrating evaluative inquiry into the organisational culture: a review and synthesis of the knowledge base. The Canadian Journal of Program Evaluation, 19(2), 99-141.

[8] Christie, C. A., & Alkin, M. C. (1999). Further reflections on evaluation misutilization. *Studies in Educational Evaluation*, 25, 1–10.

evaluation process by limiting access to data. Findings misuse includes actively misrepresenting evaluation information.[9]

Development agencies have mostly considered and tracked evaluation use through responses to recommendations. This type of use can be seen as one type of instrumental use. Using this measure, use has improved over time. A 2010 OECD study of donor evaluation systems raised issues regarding evaluation reports not being available and management recommendations not being responded. The follow-up study in 2023 showed that these issues had largely been addressed.[10][11] A range of other issues in evaluation use emerged, however, which are discussed further on page 14.

Development agencies have also considered enlightenment use to some extent. That is, users apply evaluation findings in other contexts or to other programs. For example, DFAT noted enlightenment use of evaluations in its 2014 review of evaluation use as shown in the following quote.

> *"Evaluation use covers whether direct action is taken as a result of an evaluation, and whether lessons from an evaluation are made available and applied to other investments."*[12]

However, there is a lack of discussion of the other definitions of evaluation use in the development evaluation literature. Key definitions that are missing are: conceptual use, symbolic use, and process use. Equally, there is no discussion of evaluation misuse in the development field.

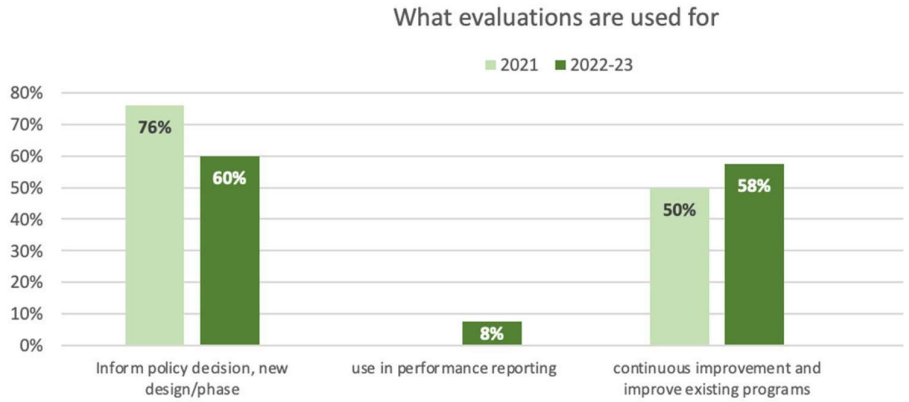
2. What we found through this review

Almost all evaluations were used (45 of 46) and one was intended for future use.

The following findings on evaluation use show that the ways DFAT used evaluations mirrors the types of use identified in the literature outlined above.

Instrumental use

Almost all the 46 evaluations were used instrumentally as shown in the table below. Most evaluations were used to inform policy decisions (for example, on extending a program to a new phase) and to improve existing programs. Some evaluations were used to report on performance.



[9] See Alkin, M. C. and King, J. A. (2017), Definitions of Evaluation Use and Misuse, Evaluation Influence, and Factors Affecting Use, *American Journal of Evaluation*, Vol. 38(3), pg. 441.

[10] OECD, (2023), Evaluation Systems in Development Co-operation 2023, https://www.oecd.org/en/publications/evaluation-systems-in-development-co-operation-2023_a255365e-en.html

[11] OECD, (2010), Evaluation in Development Agencies, Better Aid, OECD Publishing. http://dx.doi.org/10.1787/9789264094857-en

[12] Office of Development Effectiveness, (2014), Review of operational evaluations completed in 2014.

Of the 9 sampled evaluations, all of them were used instrumentally. Interviewees spoke of how they used evaluations to improve implementation, including through addressing gaps identified and applying learnings. Staff strengthened processes identified as problematic and pursued opportunities identified to achieve planned outcomes or milestones. Key areas for strengthening in subsequent program phases were gender equality and social inclusion, MEL, governance and partnership engagement as noted in the quote below.

> "*Some (findings) were used to strengthen the latter half of the program and others to inform the future phase… the report was used and the evaluation was used to test the modality itself. It has been very helpful and guided our Phase 2 design. I would say one thing was most useful was to improve the progress reporting. We did an update of the ToC and the results framework. Those changes were used for Phase 2 as well.*" Interviewee

Evaluations were used to record progress, processes and challenges and to identify lessons. They were key for maintaining institutional knowledge in the context of high staff turnover as shown in the following quote.

> *"It was really useful for me. I was new to the section. That was the case for the whole team bar one person who was working remotely. There was no one around who had corporate knowledge. For me it was my source of truth. How the program was structured, how management had changed over time."* Interviewee

Staff spoke of using evaluations to inform decision making on whether to continue or extend investments. For example, one interviewee spoke of how the evaluation "provided a good basis for identifying gaps and next steps".

We also found that evaluations were used in ways in addition to instrumental use. The process of evaluation was used, and evaluations were used conceptually, symbolically and for enlightenment.

Process use

Of the 9 sampled evaluations, DFAT used the evaluation process in 3 cases to constructively engage with the partner. In one case the evaluator helped to redefine the scope of support, after expansion due to COVID. "*We had to respond the way we did during COVID 19 which created the expectation post COVID that that would continue. We were lucky the Team Leader had a very strong relationship with the Ministry and was very respected. He just needed the evidence and justification".* Staff also spoke of learning skills in procuring and managing an evaluation. *"It (the evaluation) has made me a better M&E person overall. I have seen it done in different ways. I can think back to what worked well and what did not. I can offer suggestions. That has been great from a professional sense."*

Conceptual use

DFAT used evaluations conceptually. That is, the findings helped them think differently about the program. In one case the evaluation provided important findings on program and risk management. These findings helped the team reconsider how to manage programs when neither DFAT nor the implementing partner can be on-the-ground, particularly in conflict contexts. In another case the review of a partnership occurred in the context of a broader disaster risk reduction portfolio review. The evaluation was one way the team reconsidered support to the partner country in conjunction with a political economy analysis which was also commissioned. "*We reaffirmed some knowns, discovered some unknowns and elaborated known unknowns i.e. something was not going right but we didn't know why.*"

Symbolic use

One evaluation was used symbolically. That is, the evaluation was used as a sign of good management. DFAT used the evaluation to showcase its partnership approach to the region. "*We were not just a donor evaluating our grantee*." The collaborative approach to the evaluation modelled the way the partners worked together more broadly. Similarly, the partner wanted to demonstrate to other donors

their preferred way of working by showcasing one of their most advanced partnerships. This was intended to assist the partner in securing funds not only from other donors but also from the private sector and other diverse sources.

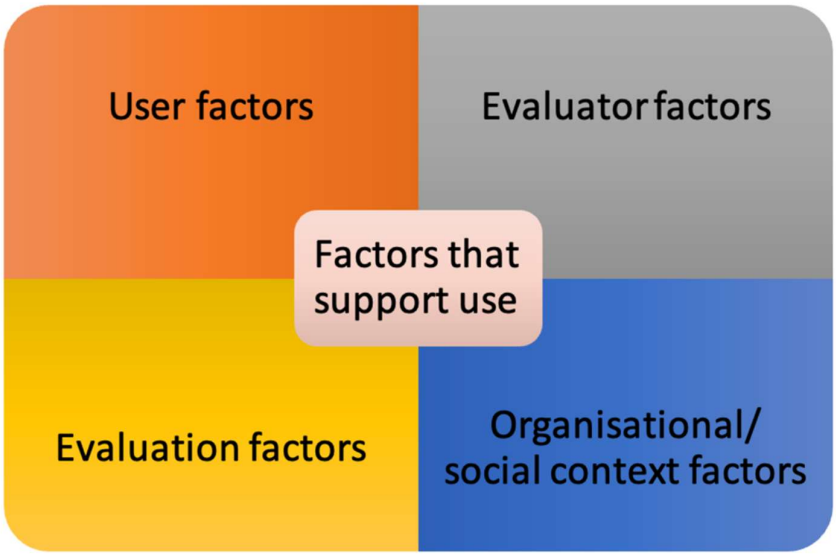
Enlightenment use

One evaluation was used for enlightenment or broader learning than just the program team. The evaluation contributed more widely to programming in the country as the team shared the design update, based heavily on the strategic review, with other development partners i.e. World Bank, to promote complementarity of investments.

Key enablers and barriers to the use of program evaluations

1. What the literature tells us

Evaluation theory provides insights into the key factors that help or hinder use. How these factors play out in DFAT evaluations will be discussed in the next section. There are 4 main categories of factors that contribute to evaluation use.[13] The 4 categories are shown in the diagram below and explained further in the following sections.[14]



User factors

User attitudes to evaluation, both in general and specific to the current evaluation, determine use. Equally, user commitment to use makes a difference. Meaningful involvement of the user in the evaluation can improve their interest and attitude.

Evaluator factors

Similarly, the commitment of the evaluator to use is critical. Also of importance, is the political sensitivity and credibility of the evaluator. Evaluators should engage potential users in the conduct of the evaluation as well as develop rapport and good working relationships to enhance evaluation use.[15]

[13] These factors were identified through 3 major empirical studies
[14] This section on factors that support use draws directly from Alkin, M.C., King, J.A., (2017), Definitions of Evaluation Use and Misuse, Evaluation Influence, and Factors Affecting Use, American Journal of Evaluation
[15] Johnson et al., 2009

Evaluation factors

Methods need to be appropriate to the evaluation question and credible to potential users. Evaluation information should meet the perceived need of users. Reporting must be in a form that users can understand and timely. Importantly, the evaluation must bear some relationship to the existing information of major users.

Organisational/ social context factors

Organisational factors are the program and larger organisational context, including how much autonomy the commissioning unit has. The age of the program and extent of its development are also important factors. Social factors include the role of other agencies, sources of information beyond the evaluation likely to be employed in decision-making, and the influence of the community on the program.

What is missing from the literature

While these categories are useful there are a few gaps in the literature. Key issues are the **lack of attention to power** and the **focus on rationality**. Nowhere in the discussion is there attention to how differential levels of power impact on the evaluation process, product or use. Evaluators and evaluation users are all navigating different power relations. These relations may relate to gender, race, or other factors and can either promote or constrain the actions of evaluators and evaluation users. Power is not static, rather continuously negotiated and in flux. As such, it is important to recognise power as an issue to be considered. Similarly, the literature only focuses on rational decision-making. However, evaluators and evaluation users do not act only rationally. The technical accuracy of a report is not the only factor that impacts on use. Rather, recent research points to the importance of emotions in decision-making.[16] How a user feels about an issue is as important as how they think. Evaluators need to pay attention to considering multiple ways evaluations can be influential, outside of logic and rationality.

Development evaluation

The factors that support and constrain evaluation use have also been considered in development evaluation. The 2016 OECD study of evaluation systems identified a range of factors seen to enhance use.[17] Evaluations were more likely to be used when there was an organisational culture and support of senior managers for that use. Well-written, timely reports and summaries that were easy to access were more likely to be used. Members were using a range of modes of dissemination including forums, conferences and workshops to increase use. While these methods were helpful, members noted that target audiences had low absorption capacity given competing priorities and volume of information available.

There are a range of focus areas identified in the 2023 OECD study of evaluation systems to increase the use of development evaluation. These are addressing the tension between providing timely advice while ensuring quality and addressing the absorption capacity of audiences to act on evaluation

[16] See for example, Lerner, J.S., Li, Y., Valdesolo, P., and Kassam, K.S., (2015), Emotion and Decision Making, The Annual Review of Psychology, 14(11), pp. 33.1-33.18

[17] OECD, (2016), Evaluation Systems in Development Co-operation, OECD Publishing. https://www.oecd-ilibrary.org/development/evaluation-systems-in-development-co-operation_9789264262065-en

findings.[18] To respond to these challenges, the 2016 OECD study showed that members are focused on the following points:

- Identifying learning objectives early in the evaluation cycle
- Engaging early with staff across the organisation to increase relevance and ownership
- Aligning evaluations with organisational objectives and ensuring timely release of findings
- Conducting more cross-cutting analyses and syntheses
- How evaluation findings are presented and disseminated

2. What we found through this review

The team examined the correlation between 5 factors and use:

- evaluation/ issue prioritisation
- evaluation process
- quality of evaluation
- systems that facilitate use
- organisational evaluation capability

We found that each of the above five factors was highly correlated with use of the sampled evaluations. For one of the factors, evaluation quality, the interviews have helped us develop a broader definition of the term. In addition to the 5 factors, the importance of the expertise of the evaluator or evaluation team emerged as another factor from the interviews. We also found that each of the factors can act as a barrier if not managed well. The section below provides further detail of findings and Annex 1 provides further detail from the interviews.

Correlation between use and evaluation/ issue prioritisation

Overall, the interviews highlighted that use was more likely where there was clarity of purpose for why the evaluation was conducted and the issues to be examined. A clear purpose aided more direct application and use of findings and recommendations. Interviews indicated that evaluation purpose is usually identified by the program area, with broader engagement to further scope and clarify key focus or issues to address, to be incorporated into the Terms of Reference. The interviews found the purpose was predominantly focused on informing strategic and programmatic decision-making. That is, evaluations were often focused on whether and how to continue investments and/or the design of subsequent investment phases, including the strengthening of key partnerships as illustrated in the quote below.

> *"We hoped to get a lot out of it. We finished the first Facility of [the program] – we needed it [the evaluation] if we wanted to do more in the development space." (Interviewee)*

Commonly evaluations were conducted as a separate activity prior to a programmatic design phase, rather than as part of an integrated evaluation/ design effort. This was generally considered among evaluation managers to be more helpful in effectively being able to apply evaluation findings. Evaluation processes seemingly gave time to share, digest and process key evaluation outcomes. Building from findings and recommendations of previous evaluations of the same program was also identified as important.

There were some key factors which appeared to either enable or hinder effective evaluation/ issue prioritisation which emerged through the interviews, survey and quality reviews which are summarised below.

[18] OECD (2023), Evaluation Systems in Development Co-operation 2023, OECD Publishing, Paris, https://doi.org/10.1787/a255365e-en

Enablers

- Initiation of planning well in advance of the evaluation activity
- Clarity of purpose of how the evaluation will be used, including both primary and secondary purposes
- Senior management engagement in the scoping and framing of the evaluation

Barriers

- Insufficient staff time dedicated to evaluation planning and delivery

Correlation between use and the evaluation process

The evaluation processes were found to be key in enabling evaluation use. Findings suggest the importance of engaging the DFAT program area engaging the evaluation team through all stages of the evaluation - from planning, procurement, implementation and report finalisation. This inclusion is likely to influence both the quality of the process and output, as well as boost staff trust in the process and findings, which ultimately encourages use. The quote below shows how DFAT staff used the evaluation process to work through issues of concern.

> *"The First Secretary appreciated the opportunity to talk through issues – [the evaluator] was like a facilitator. It was the entire interaction and dialogue around it that was helpful as it hadn't been an easy time." (Interviewee)*

Interviews suggested that where DFAT engaged well with evaluation teams, this facilitated clarity of scope and key issues to explore, consensus on approach and methodology, effective feedback processes, helpful discussion on findings and recommendations, and the development of reports more likely to have traction with their key audiences. Good engagement by DFAT of the evaluation team is also likely to boost learning in relation to evaluation processes, to the benefit of the specific evaluations and future evaluations. A facilitated engagement between DFAT and partner government counterparts through the evaluation can also enable consensus building on scope, key findings and recommendations, also key in facilitating subsequent evaluation use.

There were some enablers and barriers which appeared to influence the effectiveness of the evaluation process, which affected extent of evaluation use, summarised below.

Enablers

- DFAT intensive engagement with the evaluation team throughout the evaluation process
- Specificity and detail within the methodological approach
- Engagement with a breadth of key informants beyond implementing partners
- A facilitated engagement between DFAT and the partner government counterpart to build consensus on scope, findings and recommendations
- Tailoring and targeting in the discussion of findings with both internal and external stakeholders

Barriers

- Lack of substantial involvement or engagement of key funding and implementing partners in the design and delivery of the evaluation
- Turnover of DFAT staff, including the evaluation manager and senior managers

Correlation between use and evaluation quality

Quality was originally assessed in relation to the evaluation report.[19] Independent assessors reviewed the report against a set of 9 criteria. The rating for the criteria of quality of evidence and analysis served as the proxy for overall quality. This was because the 2012 review of evaluation quality found that this criterion most highly correlated with overall quality ratings.

What emerged from interviews, in contrast to this definition, is that staff gave emphasis to two key dimensions of quality. Firstly, a range of skills, perspectives and insights within the evaluation team which facilitates a quality product reflective of complex, local programming contexts as shown in the quote below.

> *"Having the right people was key and they had good local insight, experiences and contacts. They were able to ascertain as to whether the modality was the right one and the extent to which support provided was well aligned with key areas, then they came up with sensible and feasible recommendations which were also front of our mind."* (Interviewee)

Secondly, concise and easy-to-read reports with particular focus on high quality executive summaries.

In addition to these, the review team also emphasised a solid evidence base within reports, with specific data linked to findings and data effectively corroborated across sources as reassurance of quality. This dimension was considered less important to staff, where instead there was more inherent 'trust' in the evidence base if an evaluation team with useful skills and experience was in place. It is recognised that developing a readable evaluation report whilst providing a detailed and insightful evidence base can be a hard balance to strike. That is, evaluators must be able to communicate well while evidencing their findings.

Staff tended to have a higher perception of the quality of evaluations than the independent reviewers, assigning a comparatively higher rating for quality. Quality ratings were also assigned through the quality review process, ranging from 2-6 with an average rating of 3.6 across those sampled. Five of the nine sampled were classified by the review team as 'unsatisfactory' (rating of 3 or below).[20] In contrast, the evaluation managers assigned an average rating of 4.8/Adequate-Good (scale of 1-6) for quality across all evaluations reviewed. The difference in ratings is likely because of the familiarity of program teams with the process and the decisions made which shaped it. Perception of quality therefore appears to be linked to usage more than actual (independently assessed) quality. All evaluations were used with various uptake reported.

Key enablers and barriers of quality in evaluations are summarised below.

Enablers

- A breadth of relevant skills, perspectives and insights within the evaluation team
- Concise, readable reports
- Well-structured and effectively synthesised points within the executive summary

Barriers

- Insufficient focus given to the strategic and complex programming context
- A lack of detailed evidence as a basis for findings and recommendations

[19] A recommendation of the 2023 review was that quality should be reviewed more broadly than just the report.

[20] The evaluation team used the quality ratings from the 2023 review for the evaluations completed in 2022. Additionally, the team undertook quality reviews, using the same methodology, for the sampled evaluations completed in 2023.

Correlation between use and systems that facilitate use

The team found that over half of the systems noted in DFAT's Development Evaluation Policy are supporting evaluation use. Interviews show that three of five systems are facilitating use of evaluations.

- **Management responses:** Interviewees noted that the process of developing the management response ensured DFAT accountability to act on findings and recommendations. For example, one evaluation manager reported how developing the management response made the team consider the feasibility of recommendations "when we… were writing the management response we needed to think about the framing and whether it can be applied or not".
- **Publication:** All the sampled evaluations were published on the DFAT website which increased the opportunity for use. However, most teams reported that making sure final evaluation reports met accessibility guidelines was time consuming, difficult, and often left to DFAT staff despite being an evaluation contractor responsibility.
- **Linking design and evaluation:** Most of the sampled evaluations were designed to inform the next phase of investment, with the remainder focused on improving the current phase.

In contrast, there was less evidence of the following systems facilitating use of evaluations:

- **Evaluation findings can be easily accessed:** Several staff noted that evaluation reports can be difficult to find through the DFAT filing system. Additionally, some teams noted the often siloed nature of work teams.
- **Senior management oversight:** A minority of teams (two of nine) referred to senior management oversight. Where this was the case, it was noted as a positive. For example, in one case the Deputy Head of Mission oversaw the program and was therefore engaged in the review of the evaluation report and management response.

DFAT staff also referred to other systems that supported use, outside of those referenced in the policy. Where this was the case, there was variability in staff knowledge about these systems to facilitate evaluation use. Key here was staff turnover. Staff tended to have more knowledge on systems and processes the longer they had been in DFAT. When they were aware of such systems, staff generally reported use and appreciation of them.

The supportive systems mentioned included Performance and Quality staff engagement sessions and DFAT's intranet with internal resources available to guide evaluation related processes, such as the development of terms of reference (ToRs) and improving accessibility of documents as noted in the quote below.

> *"There is support in DFAT lunchtime sessions which focus on various topics i.e. sector wide and thematic ones, also methodologies like impact assessment. These sessions are helpful." (Interviewee)*

Specific outsourced resources, such as the Specialist Health Service (SHS) and the Quality Investment Support Services (QISS) in Indonesia, were considered key in enabling commitment of relevant and quality human resources to strengthen evaluations. The in-house procurement facilities at some large posts was also noted for enabling the hire of local expertise with good insight into the local evaluation context. All these systems were seen as key in contributing to evaluation quality which incentivised application and use of findings and recommendations.

The sharing of staff experience in conducting evaluations, even if indirect or informal, was also seen as valuable. Internal meetings and processes where findings and recommendations are discussed and decisions recorded, such as investment performance reporting processes, were seen as useful for facilitating uptake and application of evaluation findings. The lack of these processes was also noted as a constraint as shown in the quote below.

> *"What didn't really happen was more consideration of how to take the results and socialise them at different levels of the organisation. It would have been better if [the process was] structured to enable operational people to have more conversations focused on the detail coming out of the evaluations." (Interviewee)*

However, some staff also reported limited guidance around how to 'socialise' results and recommendations both internally and externally after completion of the evaluation.

Key enablers and barriers of systems that facilitate use of evaluations are summarised below.

Enablers

- Use of the following DFAT systems to support use as noted in the policy: Management responses, publication and linking design and evaluation
- Availability and knowledge of internal resources to guide evaluation-related processes
- Outsourced resources such as SHS and QISS to strengthen commitment of relevant skills and experience
- In-house procurement facility at post to boost commitment of local expertise
- Informal sharing of evaluation related experience among staff
- Internal meetings and processes focused on discussing findings of evaluations and recording decisions

Barriers

- Lack of use of the following systems as noted in the policy: Evaluation findings can be easily accessed and senior management oversight
- Silos within DFAT can limit useful sharing of evaluation findings across teams
- Lack of examples around how to 'socialise' results and recommendations, both internally and externally

Correlation between use and organisational evaluation capability

Evaluation managers considered themselves to be confident in managing evaluations, though not all staff felt confident with more technical aspects of their role commissioning and managing evaluations, such as finalising ToRs, drafting evaluation questions and reviewing evaluation plans and reports using DFAT Standards.

Staff reported a general motivation to learn more about evaluation processes and approaches. However, at the same time, staff at post reported a lack of incentive to engage usefully with evaluations, despite the availability of required financial resources. This seems to relate to a lack of time, staff expectations of limited access to relevant technical experience and expertise to enable effective delivery of evaluations, and varying perceptions in the extent of support from senior management towards evaluations which may later influence application of findings.

Key enablers and barriers of organisational evaluation capability are summarised below.

Enablers

- Evaluations seen as valuable for strengthening management, reporting, learning, partner engagement, programmatic delivery and accountability
- High levels of motivation to learn more about evaluation
- Most respondents know how to access internal standards and guidance on evaluations

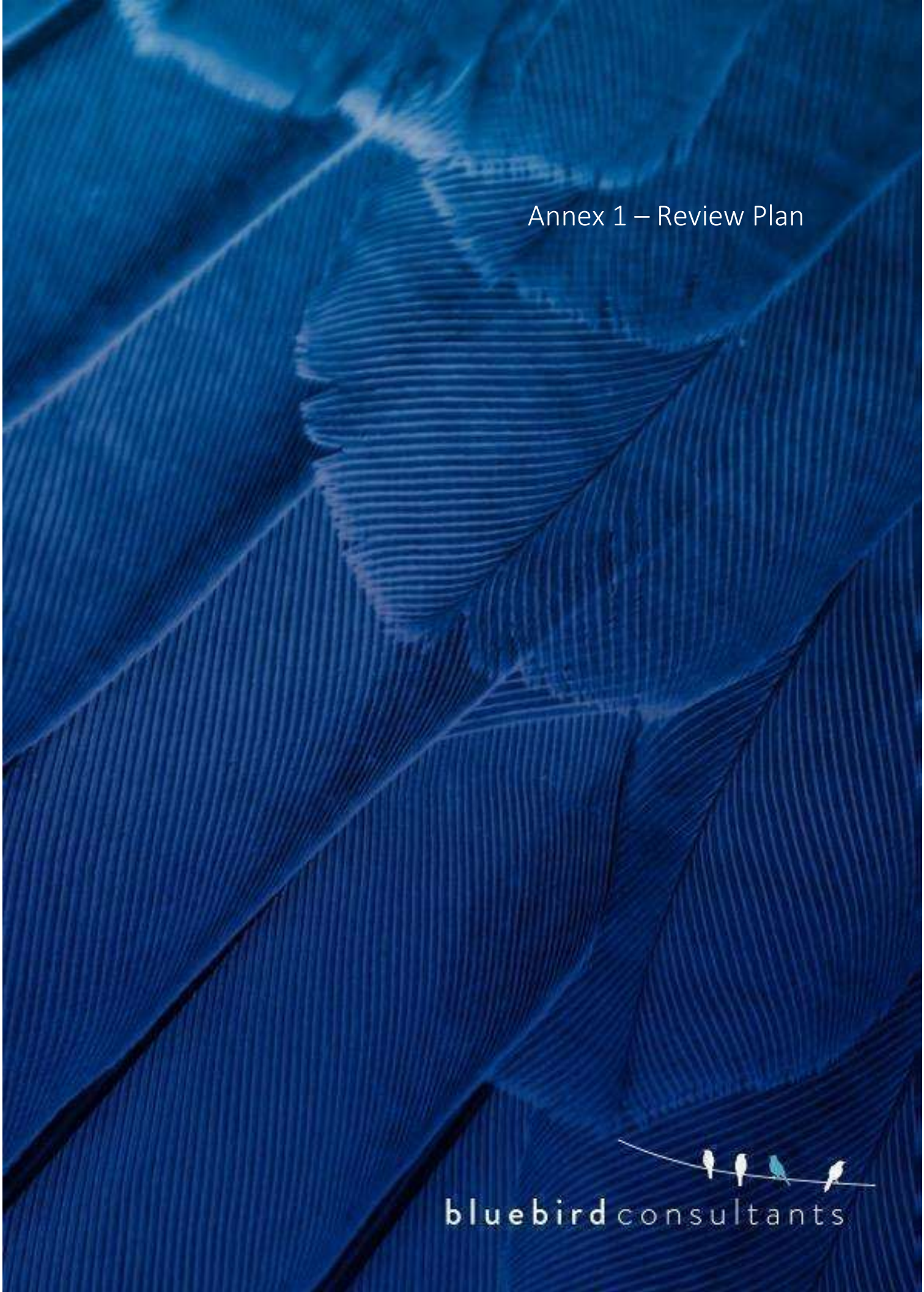
Barriers

- Varying confidence in some more technical aspects of commissioning and managing evaluations such as finalising TORs, drafting evaluation questions and use of DFAT Standards

- Varying perceptions in extent of senior management support towards evaluations and anticipated subsequent use of evidence
- A lack of time to effectively engage in evaluation activity
- Limited access to technical expertise for evaluation teams through which internal evaluation capability can be strengthened

Correlation between use and evaluator

The evaluator is a key enabling factor identified in the literature and identified as a priority in DFAT's Standards. This factor emerged as important through the interviews and was noted as a part of evaluation quality. Interviewees described how skilled evaluators brokered processes that enabled useful interaction between DFAT and partners. Skilled evaluators knew how to develop a high-quality yet accessible evaluation report and recommendations that were useful and feasible. Skilled evaluators evidenced their findings. In contrast, there were examples where evaluators did not meet the mark. Some evaluators were unable to link different sets of data coherently into a broader analysis. Other evaluators wrote reports that required significant input from DFAT officers to meet DFAT Standards. One factor that stood out from interviews was that DFAT teams seemed to have inherent trust in the evaluation team based on their skills and experience and therefore paid less attention to the actual quality of their work.



Annex 1 – Review Plan

bluebird consultants

2024 Review of the quality and use of DFAT 2022-2023 evaluations:

Review plan

1. Introduction

The Development Evaluation and Assurance Section (EVS) of DFAT and Bluebird Consultants will conduct a review of the quality and use of development evaluations completed in 2022-23 (the review). This work follows four previous reviews of program evaluations, which examined evaluations completed in 2012, 2014, 2017 and 2022. This document outlines the proposed method and budget for the review to take place in 2024.

2. Review Questions

The review will answer the following evaluation questions.

1. How are program evaluations used?
2. What are the key enablers and barriers to the use of program evaluations? Specifically examining correlation between use and:
    a. Evaluation/ issue prioritisation
    b. Evaluation process
    c. Evaluation quality
    d. Systems that facilitate use
    e. Organisational evaluation capability
3. What actions are recommended to ensure better use of program evaluations?

3. Methodology

The review will be conducted by a team of three comprising two external consultants and an EVS staff member. The external Team Leader is from the EIS Support core team and the second external team member will be sourced from the Help Desk. The Team Leader was involved in the 2022 review of quality and so will help to bring consistency to this review.

The review will be conducted over three stages.

Stage 1: Literature Review, Quality Assessment and Use Survey (all evaluations in scope)

The initial stage will involve a literature review and collecting preliminary data on evaluation use and quality from the full cohort of 46 evaluations and evaluation managers.

The team will examine the **evaluation literature on use** which will inform the data collection and analysis phases. A preliminary review of the literature shows the most common understanding of use is using evaluations to improve programs by providing information for action and decision-making.[21] The literature describes this type of use as instrumental and we will use the following terms that are relevant to DFAT in the review: 'policy development', 'program design', 'performance reporting', 'program improvement', and 'thematic learning'. Empirical research documents other ways that evaluations are also used:

[21] Weiss, C. H. (1998). Have we learned anything new about the use of evaluation? American Journal of Evaluation, 19(1), 13.

- *Conceptually* - to help program staff understand a program in a new way. We will refer to this type of use as 'learning'.
- *Symbolically* - as an instrument of persuasion or to legitimate an existing position. We will refer to this type of use as 'justification'.
- *Enlightenment* - where evaluation findings added knowledge to the field and were used broadly. We will refer to this type of use as 'contributing to the field'.
- *Process use* – or involving users in the process of evaluation – was a key use of evaluations and also a strategy to enable other types of use.[22] We will refer to this type of use as 'learning from evaluation'.

The team will draw from the 2022 desk review that examined quality and the 2022 survey to examine use. The team will also include the 6 evaluations from 2023 for quality and administer a survey on use to these program teams. The team will review this preliminary data to identify a purposeful qualitative sample of 9 cases for deeper analysis through interview. The selection will consider the use (and/or acceptance and relevance) of recommendations as noted below.

- **Low:** Some recommendations not accepted or partially accepted
- **Moderate:** Most recommendations accepted
- **High:** All recommendations accepted

Where feasible the team will also seek representation across the following categories:

- geography
- size of Posts
- partner managed vs DFAT managed
- thematic areas
- types of implementation partner

Stage 2: Deeper Insights. Survey on evaluation capability and interviews for the 9 Cases

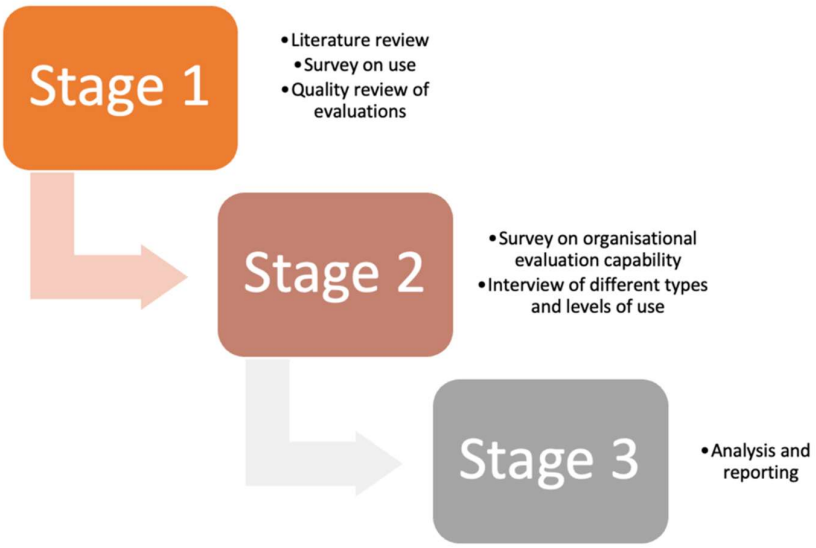
The second stage will involve the review team surveying and interviewing key stakeholders from each of the selected evaluations. The team will collect data through a survey of the evaluation manager on organisational evaluation capability. The team will then interview the evaluation manager, program manager, senior manager and evaluator on key enablers and constraints, specifically examining the:

- importance of prioritisation of key issues or programs
- evaluation process
- quality of the evaluation report
- systems that facilitate use

Stage 3: Analysis and Reporting

The third and final stage will involve the team analysing the data to identify key findings and provide examples of different types and levels of use. The team will use the Qualitative Comparative Analysis (QCA) method to score the presence and absence of predicted factors, whilst also examining the qualitative detail of each case of use.[23] The team will collate preliminary findings and workshop these with the EVS and EIS support core team. The review team will then finalise the report, including developing a short snapshot of findings for executives, and communicate findings. Guided by EVS, the team could communicate findings at a DFAT learning event and in other relevant internal and external forums.

[22] Patton, M. Q. (1998). Discovering process use. Evaluation, 4(2), 9.
[23] For further information see: Simister, N. and Scholz, V., Qualitative Comparative Analysis, INTRAC for Civil Society, https://www.intrac.org/wpcms/wp-content/uploads/2017/01/Qualitative-comparative-analysis.pdf



Detail on how each evaluation questions will be answered is below.

Q1: How are program evaluations used?

This question will be answered by collecting data through both a survey of all program managers and interviews with a sample of evaluations. The team will draw from the data collected by survey from the 2022 cohort. The 2023 use survey will include minor modifications to the 2022 survey to capture the range of types of uses outlined above.

Survey of Evaluation Use (Excel spreadsheet data table summary)

- Region
- Country/ Regional Program
- Evaluation Title
- Evaluation type
- Investment sector
- Mgt Response No. of recommendations accepted
- Mgt Response No. of recommendations partially accepted
- Mgt Response No. of recommendations not accepted
- Qualitative Q1. How have agreed recommendations been used to improve programs or inform future programming? (eg inform policy decision or new design, use in performance reporting, continuous improvement, improve existing programs, thematic learning)
    - How have evaluations been used (coded categories)
        - 1 – Inform policy decision
        - 2 – Inform program design/phase
        - 3 – Use in performance reporting
        - 4 – Improve existing programs
        - 5 – Staff learning
        - 6 – Thematic learning
        - 7 – Contributing to the field
        - 8 – Justifying an existing position
        - 9 - Compliance
        - 10 – No use

        - 11 – Other
    - Areas that Evaluations have improved (free coding)
        - 1. GEDSI
        - 2. M&E
        - 3. Relevance (eg respond to and address new strategic and policy priorities)
        - 4. Efficiency (budget and funding decisions, governance, resourcing, expertise, harmonisation, consolidation)
        - 5. Risk and safeguards
        - 6. Sustainability
        - 7. Thematic and technical areas
- Qualitative Question Q2. If any agreed recommendations haven't yet been implemented, identify the factors that have hindered the uptake of the recommendations
    - Barriers to implementing recommendations (free coding)
        - 1. Capacity, resources and commitment (partner, partner government or DFAT)
        - 2. To be implemented in new phase or program, or by another team
        - 3. Being addressed through another mechanism, eg high level partnership consultations, other committee/governance structures
        - 4. Not in line with DFAT/Australian government policies/requirements
        - 5. Waiting for direction from new International Development Policy
        - 6. No longer relevant – insufficient time remaining in program to implement, changed context, program not continuing
- Qualitative Question Q3. Please share any other important issues related to use not noted above.

Interviews on evaluation use

The team will interview a sample of key stakeholders on how evaluations were used. The team will interview a maximum of 2 stakeholders per evaluation. Stakeholders may include the Evaluation Manager, Investment Manager, Senior Manager of the Section and the Evaluator.

Analysis

The team will analyse the survey data and code evaluations for levels of use: high, moderate and low. The team will code the interview transcripts thematically and identify key themes related to use.

The interview sheet will include a section where evaluation use can be identified as good practice. These will be reviewed by the team leader. In addition, the top-rated evaluations will also be considered by the team leader. A limited number of good practice examples with a basic description will be proposed to DFAT based on this process. The team will then develop cases of good levels and types of use. We will also develop a generic list of factors that impede evaluations being used drawing from the cases of poor practice. Lessons from poor practice will be shared for internal use and not include identifying details.

Q2: What are the key enablers and constraints to the use of program evaluations?

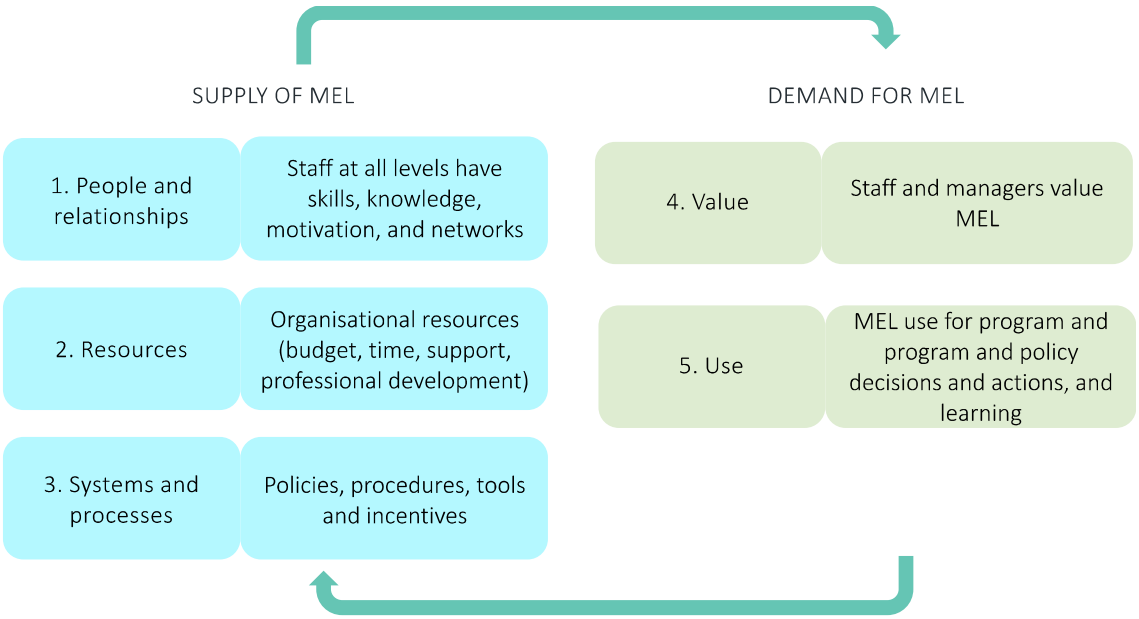
This question will be answered by collecting data through both a survey of all sampled program managers and interviews with a sample of evaluations. The evaluations selected in the sample will represent different use cases: 4 cases of high use, 4 cases of moderate use and 4 cases of low use as defined below.

Survey of organisational evaluation capability

- **Evaluation managers will be surveyed on the organisational evaluation capability of their post/ unit.** The survey will explore the 5 domains of organisational evaluation capability as identified in the empirical and theoretical literature on evaluation capacity building.[24] The 5 domains are:

[24] See Cousins, J. B. & Bourgeois, I. (Eds.), (2014), Organizational capacity to do and use evaluation. New Directions for Evaluation, 141, San Francisco, CA: Jossey-Bass and Preskill, H., & Torres, R. T. (1999a). Building capacity for

people, resources, systems, value and use. The concept of organisational evaluation capability and each domain is further explained in the diagram below.



- In relation to the people domain, the survey will draw from the 4 dimensions of capability of evaluation managers used in the UK Government's Analytic Evaluation Capabilities Framework. These 4 dimensions relate to:
    - Scoping an Evaluation
    - Leading and Managing an Evaluation
    - Methods of Data Collection and Analysis
    - Use and Dissemination

Survey of organisational evaluation capability

- As noted above, the team will interview key stakeholders on a sample of evaluations representing high, moderate and low use cases. The interview will explore four key areas as identified in the table below. The section below shows the interviewees most likely to provide information on each key area.
    - **Priorisation of the evaluation issue or program.** DFAT's evaluation policy notes that evaluation resources will be prioritised to the most important issues or programs. The interview will examine the extent to which the questions considered by the evaluation were prioritised issues of evaluation users. Equally, the interview will examine whether the program itself was prioritised for evaluation.
        - Senior manager
    - **The evaluation process.** The interview will explore the way in which the evaluation was conducted and the extent to which DFAT was involved. Evaluation literature identifies engagement in the evaluation process as a type of use which also can lead to higher levels of other types of use.
        - Evaluation manager
        - Evaluator

organisational learning through evaluative enquiry. Evaluation, 5(1), 19 and Milne, C., Bowen, B., Merlene, M., (2015), Getting serious about evaluation capacity building: Lessons from a large government agency that did it strategically then evaluated it., ARTC Consultants, AES Conference Presentation, https://www.artd.com.au/wp-content/uploads/2020/09/Evaluation-Capacity-Building-AES-2015.pdf

- **Quality of the evaluation.** The interview will examine perceptions of quality of the evaluation and compare these to the findings of expert evaluators. The interview will examine whether there is a minimum level of quality required for evaluations to be used well.
    - Senior manager
    - Evaluation manager
    - Program manager
    - Evaluator
- **Systems that facilitate use.** DFAT's evaluation policy notes the importance of internal systems to facilitate use. We will examine these systems (for example, the management response, linking evaluation and design of programs, and senior management oversight) and the extent to which these systems supported use.
    - Senior manager
    - Program manager

Using the data collected above, the team will undertake QCA to examine any relationship between evaluation use and: prioritisation, process, quality and organisational evaluation capability. The team will also look for other possible factors contributing to or constraining evaluation use.

Q3: What actions are recommended to ensure better use of program evaluations?

The team will develop recommendations drawing from the data and discussion with the broader EVS and Bluebird teams at a collective workshop held at the end of the analysis phase.

4. Team roles

The review will be undertaken by the following team members. The roles and responsibilities are noted below.

Team Leader (Bluebird)

- Draft the Review Plan and methodology for the review, update the template and manual, draft the interview protocol
- Oversee review process, lead the team, troubleshoot
- Moderate consistency of assessments across the team
- Develop survey instruments
- Conduct interviews on evaluation use
- Conduct data analysis
- Draft and finalise the report
- Liaise with DFAT

Team Members (EVS and Help Desk Consultant)

- Contribute to the Review Plan
- Assess the quality of program evaluations (up to 3 evaluations) by completing the assessment template for each assigned evaluation
- Conduct interviews on evaluation use
- Conduct data analysis (quantitative and qualitative)
- Participate in moderation and weekly meetings and final workshop to share insights and help validate findings and propose recommendations
- Input into the draft and final report

Bluebird will ensure the quality of the final report through reviews by the Bluebird Director and the Team Leader for the EIS Support Services. The final report will be copy edited and well-presented through the inputs of an editor and graphic designer. The report will meet accessibility requirements to be published on the DFAT website.

5. Limitations and risks

i. Consistency of assessments across the team and across years

The quality of program evaluation reports will be assessed by team members using the assessment template. To ensure the findings of the review are credible, it will be important to ensure team members assess program evaluations relatively consistently, and that this assessment can be reasonably compared to assessments of 2017, 2014 and 2012 evaluations. Consistency across years is a risk, but the use of the same nine criteria and DFAT Design, Monitoring and Evaluation Standards reduces this risk. Also, the team leader's involvement in the 2022 review helps minimise the risk. Consistency of ratings across the team will be maximised by having a smaller team than in prior years and through the moderation processes described above.

ii. Perceptions undermining independence

Each team member will be requested to identify any actual or perceived conflicts of interest regarding the 46 evaluations ahead of work commencing. The team leader will document these conflicts or potential conflicts and ensure that team members are not allocated the relevant evaluations to assess. Team members should raise a perceived conflict of interest at any time it becomes apparent. The two unpublished evaluations from 2022 were assigned to the DFAT staff member of the team and will be dealt with by the DFAT team member if selected in this year's interview sample.

The involvement of a DFAT staff person on the team might be perceived to undermine the independence of the quality review. This risk is reduced given the staff member is from EVS rather than a program manager. Additionally, the risk of biased assessment will be further managed through the moderation processes described above.

iii. Access and recall

There are risks related to the ability to access Evaluation, Investment and Senior Managers for evaluations that are nearly 2 years old now. Some of these staff will have moved to other positions or posts. Additionally, there will be risks related to how much these staff members recall about the process.

Working closely with EVS helps to address the risk of access. EVS can support the identification of staff who have moved to other positions but remain within the agency. There is no mitigation strategy that can be used in relation to the issue of recall. Rather, that will remain a limitation on the findings and will be listed as such in the final report.

iv. Managing sensitive information

The proposed approach raises potential risks and ethical challenges as team members may gain access to sensitive information through the document review and interview process. All team members will abide by DFAT's Ethical Research Guidelines and as professional evaluators are bound by the Australian Evaluation Society's Code of Ethical Conduct and Guidelines for the Ethical Conduct of Evaluations. All Bluebird team members have signed and are bound by confidentiality agreements.

v. Applicability of findings

There is a risk that findings gained only through the purposeful qualitative sample are generalised to all evaluations. The final report will be careful in presenting findings. The report will present findings that are triangulated across the literature review, survey and interviews.

6. Schedule

The review will take place from June to September 2024, with the absolute deadline for the final report being Friday 27 September 2024. The section below indicates the main tasks, persons responsible and approximate timeframe for the review tasks.

Timeline, Review Tasks, and Responsibilities

May

- Identify suitable team member from the Helpdesk, Team Leader

June

- Obtain declarations of conflict of interest, Team Leader
- Inception meeting to agree review plan, templates, handbook, and final report format, Team Leader and DFAT
- Administer evaluation use survey, DFAT
- Conduct literature review, Team Leader

July

- Main moderation meeting, Team Leader and EVS team member
- Assess the quality of each of remaining 7 program evaluations and record in template, Review Team members
- Identify interview sample and arrange interviews, Review Team members

August and September

- Conduct interviews, Review Team

October

- Draft preliminary findings, Review Team

November

- Provide snapshots to DFAT, Team Leader
- Team workshop, including DFAT, to discuss analysis and agree on key findings and recommendations, All team members plus DFAT
- Draft report, Team Leader
- DFAT review of report, DFAT
- Finalise report and 1-2 page summary of findings, Team Leader
- Presentation of findings at a DFAT learning event, Team Leader

7. Outputs

Outputs will include:

- This review Plan outlining the detailed methods to be used for the review, including Assessment Template and Handbook and draft interview protocol.
- A concise report outlining the key findings of the use and quality review and recommendations, including:
    - a summary (such as small set of slides or a 2-page summary) for broadly communicating the findings
    - A list of 4-6 good practice evaluation products (annexed to the report)
    - Aggregated and de-identified lessons from poor practice use examples
    - Recommendations for future Annual Quality and Use Reviews
- Between 6 and 12 ‘evaluation use’ snapshots for the PAD-C (format to be agreed)
- Detailed records of data collected.

Annex: Quality review

The quality of evaluations has been assessed using the same nine quality criteria used in the previous reviews, which are the key elements across relevant standards from DFAT's Monitoring and Evaluation Standards (Standard 8 – Independent Evaluation TOR, Standard 9 – Independent Evaluation Plan and Standard 10 – Independent Evaluation Report). The nine quality criteria are:

1. quality of executive summary
2. purpose of evaluation
3. scope of evaluation
4. appropriateness of the methodology and use of sources
5. adequacy and use of M&E
6. context of the initiative
7. evaluation questions
8. credibility of evidence and analysis
9. quality of recommendations

The review team will draw from the assessments already conducted on the 2022 cohort. The team will take a similar approach to the 6 evaluations to be assessed from 2023 (to identify a suitable case study). The team's assessment will be guided by a manual which includes greater detail on each of the nine criteria, derived from DFAT's Monitoring and Evaluation Standards. The team members include both internal (EVS representative) and external (expert evaluator/s). This approach of pairing an expert evaluator with an EVS representative ensures capability building within the Unit. Equally, it means that the review will be informed by an inside understanding of key issues and priorities and that evaluations exempted from publication can be assessed. The team will arrive at a rating of 1 to 6 (see points below) and record a narrative comment in their rationale for assigning that rating, in the assessment template.

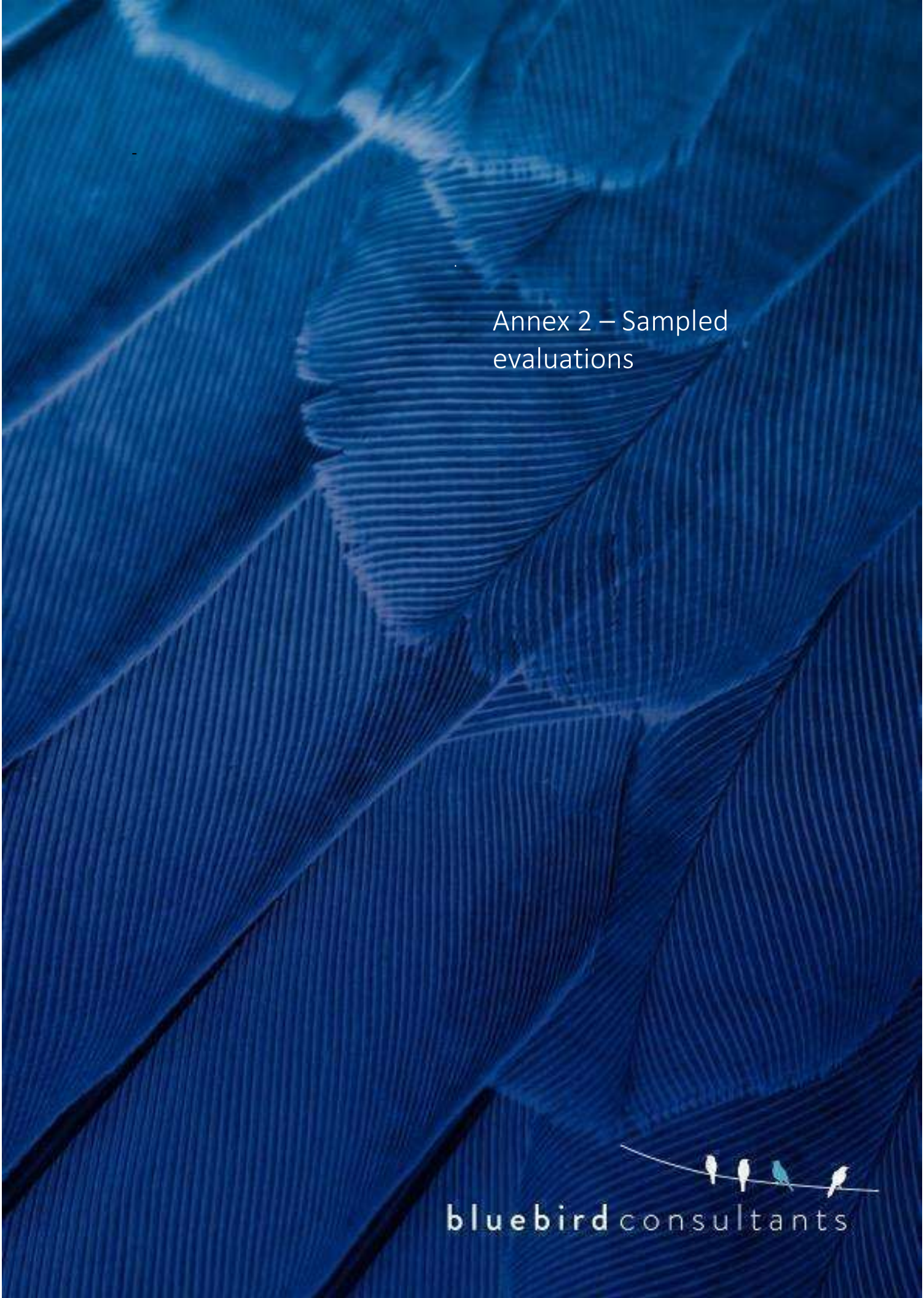
Ratings

- **6 - Very high quality:** satisfies criteria in all areas
- **5 - Good quality:** satisfies criteria in almost all areas
- **4 – Adequate quality:** on balance satisfies criteria; does not fail in any major area
- **3 – Less than adequate quality:** on balance does not satisfy criteria and/or fails in at least one major area
- **2 – Poor quality:** does not satisfy criteria in several major areas
- **1 – Very poor quality:** does not satisfy criteria in any major area

Moderation process

The ratings will be moderated multiple ways, to increase the consistency of ratings across team members and across quality reviews of evaluation reports. First the team leader was a member of the team from the review of 2022 evaluations, and this provides a level of continuity that is helpful in setting the bar for expectations of quality. Second, a moderation exercise will be conducted at the commencement of work. This will ensure the team members are familiar with the template criteria, handbook and ratings criteria, ensure language, interpretations and relative priorities of aspects of the nine criteria are well understood across the team members. This moderation exercise will establish the basis for quality assessments.

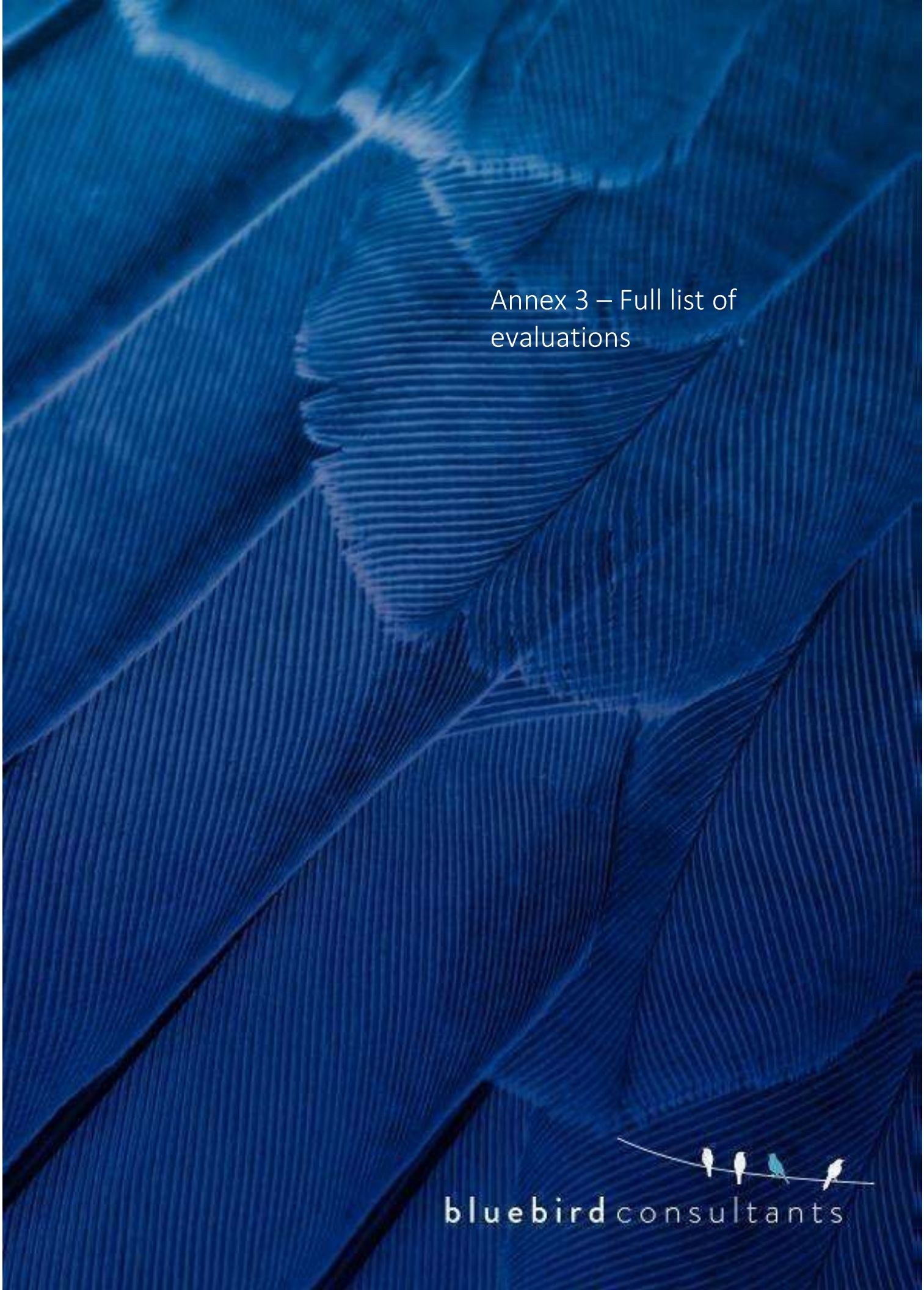
Annex 2 – Sampled evaluations



bluebird consultants

The sample comprised nine evaluations, including one exempt from publication.

Year	Evaluation title	Level of acceptance of recs.	Regional/ Bilateral	Country	Sectors	Type of use noted in initial survey
2022	End of Investment Evaluation: Pacific IUU Fishing	High	Regional	Pacific Regional	Agriculture, Trade and other Production Sectors	Design
2022	Australia-World Bank Indonesia Partnership Independent Mid-Term Review	High	Bilateral	Indonesia	Governance	Program improvement and performance reporting
2023	Independent Evaluation of the Sri Lanka Support Unit (SLSU)	High	Bilateral	Sri Lanka	Governance	Multiple
2022	Australia-SPC Partnership Evaluation	Medium	Regional	Pacific Regional	Multisector & general support	Design
2022	Australia-Cambodia Cooperation for Equitable Sustainable Services (ACCESS) End of program evaluation	Medium	Bilateral	Cambodia	Governance	Design
2023	Saferkids PH Mid-Term Review	Medium	Bilateral	Philippines	Governance	Multiple
2022	Strategic Review of the Fiji Health Program	Low	Bilateral	Fiji	Education	Design
2023	Technical Disaster Risk Reduction Program in PNG Evaluation	Low	Bilateral	PNG	Humanitarian	Design



Annex 3 – Full list of evaluations

bluebird consultants

2022 Evaluations

In 2022, 40 evaluations were completed and 38 published on the DFAT website with a management response.

Region/global/sector	Country/ Regional Program	Evaluation Title	Investment sector
Pacific	PNG	Final Review of PNG-Australia Governance Partnership	Governance
Pacific	PNG	Education Emergency Response and Recovery Plan Independent Evaluation	Education
Pacific	PNG	Independent Review of South Fly Resilience Plan	Governance
Pacific	PNG	Review of the Markets, Economic Recovery, and Inclusion Program (Phase One)	Governance
Pacific	PNG	Review of the PNG-Australia Transport Sector Support Program Phase 2 (TSSP2)	Economic infrastructure and services
Pacific	PNG	Australia, New Zealand, International Finance Corporation: Papua New Guinea Partnership Midterm Evaluation	Agriculture, Trade and other Production Sectors
Pacific	PNG	Justice Services and Stability for Development Program (JSS4D) Mid-Term Review	Governance
Pacific	Pacific Regional	Australia-SPC Partnership Evaluation	Multisector and general support
Pacific	Pacific Regional	Australian Infrastructure Financing Facility for the Pacific Two-Year System-Wide Review	Economic infrastructure and services
Pacific	Pacific Regional	Australia's COVID-19 Response Package for the Pacific and Timor-Leste Independent Review 2020-2022	Governance
Pacific	Pacific Regional	Mid-Term Review Report of the Pacific Insurance and Climate Adaptation Programme	Governance

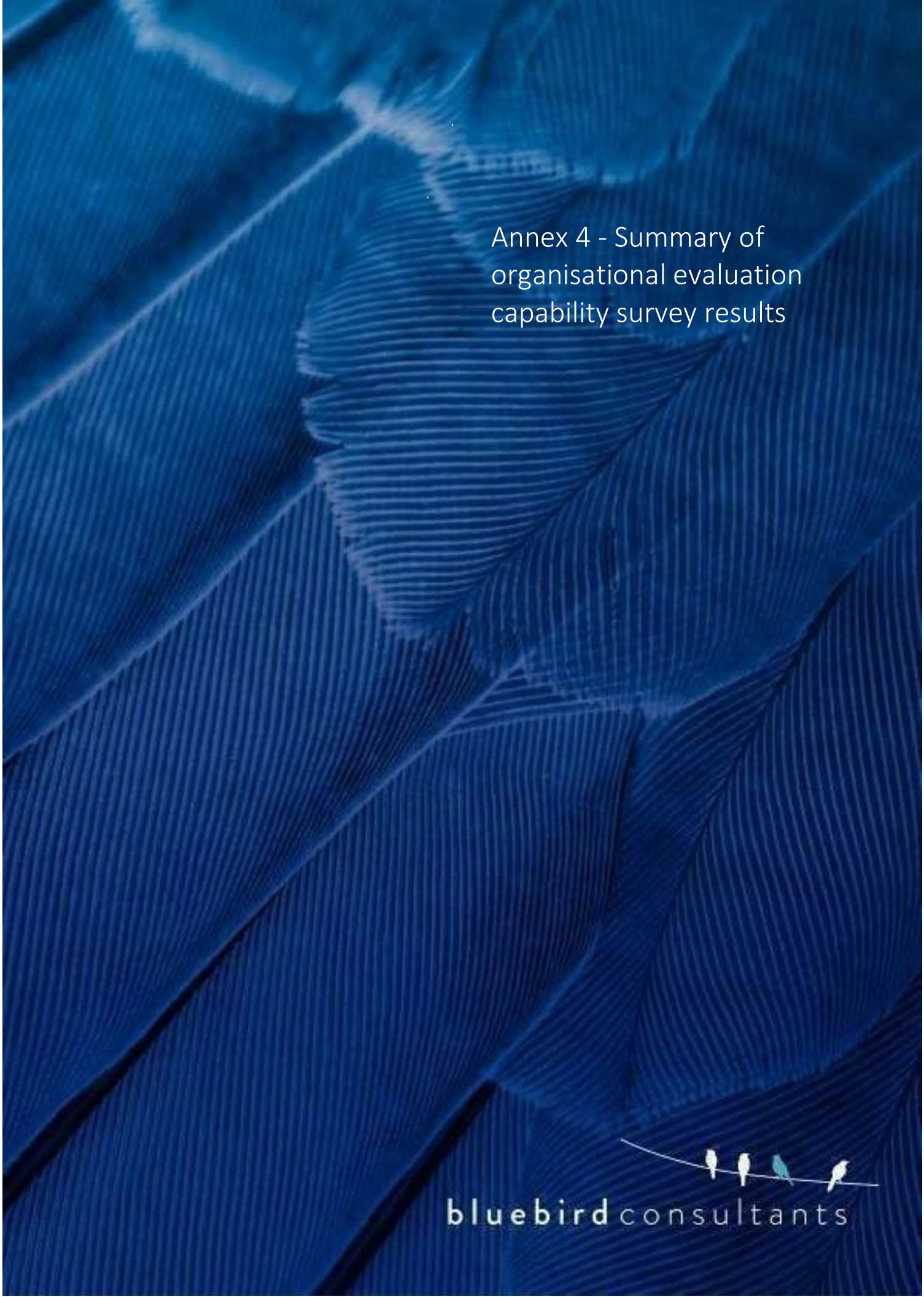
Pacific	Pacific Regional	End of Investment Evaluation: Pacific IUU Fishing	Agriculture, Trade and other Production Sectors
Pacific	Pacific Regional	Pacific Digital Economy Programme Mid-Term Review	Governance
Pacific	Solomon Islands	Ombudsman Twinning Support Independent Review	Governance
Pacific	Solomon Islands	Review and Evaluation of the Performance of Sustainable Transport Infrastructure Improvement Program (STIIP) and the National Transport Fund (NTF) in the Solomon Islands	Economic infrastructure and services
Pacific	Fiji	Strategic Review of the Fiji Health Program	Education
Pacific	Nauru	Every Life Matters: Review of DFAT Health Investments to Nauru	Health
Southeast Asia	Indonesia	Independent Strategic Review of Innovation for Indonesia's School Children Phase 2 and Rural and Remote Education Initiative for Papua Provinces Phase 3	Education
Southeast Asia	Indonesia	Penyediaan Air Minum dan Sanitasi Berbasis Masyarakat (PAMSIMAS) Final Independent Evaluation	Health
Southeast Asia	Indonesia	Australia-World Bank Indonesia Partnership Independent Mid-Term Review	Governance
Southeast Asia	Timor-Leste	Joint Independent Evaluation - Timor-Leste Police Development Program	Governance
Southeast Asia	Timor-Leste	Partnership for Human Development Mid-Term Review	Governance
Southeast Asia	Vietnam	Aus4Reform Review	Governance
Southeast Asia	Vietnam	Aus4Innovation Mid-term Review	Multisector and general support
Southeast Asia	Cambodia	Australia-Cambodia Cooperation for Equitable Sustainable Services (ACCESS) End of program evaluation	Governance
Southeast Asia	Cambodia	Ponlok Chomnes Independent Strategic Review	Education

Southeast Asia	Laos	BEQUAL Phase 1 Independent End of Program Review	Education
Southeast Asia	ASEAN and Mekong	ASEAN-Australia Digital Trade Standards Initiative Mid-Term Review	Economic infrastructure and services
Southeast Asia	ASEAN and Mekong	Mid-Term Review: ASEAN Australia Smart Cities Trust Fund (AASCTF)	Economic infrastructure and services
Southeast Asia	ASEAN and Mekong	ASEAN-Australia Counter-Trafficking Program Mid-Term Review (MTR)	Governance
South and Central Asia	Bangladesh	Program Completion Review of the Strategic Partnership Arrangement (SPA) Phase 2 in Bangladesh between DFAT, FCDO and BRAC	Education
South and Central Asia	Sri Lanka	Independent Evaluation of Women in Work (WIW) Program, Sri Lanka	Governance
South and Central Asia	Mongolia	Australia Mongolia Extractives Program (AMEP) II Mid - Term Review	Agriculture, Trade and other Production Sectors
Humanitarian	Humanitarian	Australia Assists End of Program Evaluation	Humanitarian
Humanitarian	Humanitarian	Review of the Humanitarian Logistics Capability	Humanitarian
Global	Global NGO	Independent Evaluation of the Australian NGO Cooperation Program (ANCP)	All
Sector	Health	Evaluation and Forward Scoping for the Therapeutic Goods Administration's Regulatory Strengthening Program and the Australian Expert Technical Assistance Program- Regulatory Support and Safety Monitoring	Health
Sector	Education	Mid-term evaluation of the Global Education Monitoring (GEM) Centre Phase 3	Education

January – June 2023 Evaluations

In the first six months of 2023, six evaluations were completed and published on the DFAT website with a management response

Region/global/sector	Country/ Regional Program	Evaluation Title
Pacific	PNG	Technical Disaster Risk Reduction Program in PNG Evaluation
Pacific	Vanuatu	Mid-Term Evaluation Vanuatu Education Support Program
Southeast Asia	Indonesia	Strategic Review: Australia-Indonesia Partnership for Rural Incomes through Support for Markets in Agriculture, Phase 2 (PRISMA-2)
Southeast Asia	Philippines	Australia-World Bank Growth and Prosperity in the Philippines (AGaP) Trust Fund Independent Midterm Review
Southeast Asia	Philippines	Saferkids PH Mid-Term Review
South and Central Asia	Sri Lanka	Independent Evaluation of the Sri Lanka Support Unit (SLSU)



Annex 4 - Summary of organisational evaluation capability survey results

bluebird consultants

The team administered a survey focused on organisational evaluation capability to the manager of each of the programs evaluated. Almost all (8 of 9) responded. Below are the summary results of the survey.

- Respondents value evaluations and have high levels of motivation to learn more about evaluation
- Respondents are confident managing evaluations
- However, respondents are not confident with managing technical aspects of the evaluation such as finalising TORs and drafting evaluation questions
- Most respondents are not confident to review evaluation outputs using the DFAT Standards
- There is some disagreement with statements of senior management support for evaluations and use of evidence. Some agreed strongly with these statements whereas others did not.
- Respondents say that at post they lack time and incentives to engage in evaluation but have budget
- Respondents noted that access to technical specialists can be problematic
- Most respondents say that at post evaluations are used for internal reporting, informing the next stage of a program and compliance
- Some respondents say that at post evaluations are used for learning, contributing to the field and influencing partners
- Most respondents know how to access DFAT Standards and guidance