Evaluation purpose and methodology

This section builds on the final Evaluation Plan submitted by Tetra Tech International Development to the Department of Foreign Affairs and Trade (DFAT) on 10 December 2020, to provide an overview of the evaluation purpose, key evaluation questions and the methodology used.

Purpose and use of the evaluation

The aim of this evaluation was to serve DFAT's dual key purpose: to account for Australia's investment through the program, and to improve what future investment can achieve. To achieve this, the evaluation was backward looking to provide evidence of the outcomes achieved by the program but also forward looking to provide recommendations on the policy and strategic foresight of the program by exploring what works best for whom in what contexts, as well as new approaches and ways of working/volunteering.

Evaluation use

The evaluation's Terms of Reference identified three primary uses of the evaluation findings, namely:

- Informing decisions about whether the program and the term of the current service provider/contractor should be extended by the second term option of five years
- Demonstrating program achievements to Australia's partners, stakeholders, and taxpayers
- Informing future design for the program particularly in a post-COVID-19 environment.

In addition to the primary uses above, the evaluation also provides other benefits and serves several other secondary uses. It also:

- Generates a solid evidence-base about what works – to share lessons and "best practice", informing future programming decisions
- Builds awareness and engagement with the program, its complementarity to other DFAT programs, and strengthening partner understanding of its strategic intentions.

Evaluation scope

This evaluation covered the program's activities over the period from July 2017 to 2020 including the transition phase. During this period, the program moved to a new design and contractor. This new chapter continues to deliver on the enduring vision and goal of the program to match skilled Australians with organisations overseas to deliver on their own locally determined objectives. The program is building on successful past iterations and is guided by a Global Program Strategy.

Specifically, the evaluation focussed on volunteering assignments across all 26 countries, cognisant that the Pacific Program is comparatively more substantial in terms of numbers of volunteers and assignments.

As such, the evaluation aimed to:

- Focus on the **strategic level** when looking at all activities within the 26 countries of operation
- Focus on **certain types of impacts** from the program's three key impact areas as well as key sectors (health, education, etc.) that will offer an opportunity to examine specific strands of interest e.g. different sectoral impacts and partnership structures
- Focus on **certain types of assignments** that will offer an opportunity to examine specific strands of interest e.g. different volunteering modalities
- Seek to ensure that a **wide representation** of voices and opinions are gathered, particularly local populations
- Undertake **high-level analysis of the program** to provide insights into the efficiency of governance arrangements and program management systems
- Undertake six **country-level case studies** including a range of county-level stakeholder opinions and examining two partner organisations per country.

Key evaluation questions

The terms of reference identify key questions, organised under the evaluation objectives, to assess **relevance, effectiveness, efficiency, sustainability, and future direction for the program**. The first four of these serves the evaluation's accountability purpose. The last is specifically to serve the evaluation's improvement purpose.

Below is the summary of evaluation objectives, description, and main evaluation questions.

Objective	Description	Key evaluation questions
Objective 1: Relevance	To examine the extent to which the program's design remains relevant	• How well does the design of the program respond to and continue to remain relevant to the strategic objectives of Australia's development assistance program and country strategies? • To what extent are the program's activities and outcomes aligned with the needs and priorities of partner organisations? • Is the program's logic 'fit for purpose' and being used to inform Australian Volunteers International's management of the program?
Objective 2: Effectiveness	To examine the extent to which the Volunteers Program is making progress towards its three End of Program Outcomes (EoPOs)	• Have partner organisations sustainably strengthened their capacity? • To what extent does the Australian public better appreciate the value of international volunteering? • How have volunteers gained professionally and personally? • Has the change to unified branding under the Australian Aid banner delivered greater recognition for the program? • What do stakeholders' perceptions tell us about the program's contribution to Australia's soft power?
Objective 3: Efficiency	To review the efficiency of the governance model and program management arrangements.	• To what extent does the program represent value for money? • How does the program's costs compare to those of other donors? • To what extent do corporate governance arrangements support efficient decision making and program management?
Objective 4: Future direction	To inform future directions of the program	• To what extent has the program been affected by COVID-19 and other issues? • How is the program adapting to its changing external environment? • What do stakeholders perceive as the key challenges and risks facing the program in the future? • What strategies, alternative modalities and approaches could be considered to strengthen the program's resilience and sustainability into the future?

To ensure there is awareness of the evaluation and to obtain buy-in from DFAT stakeholders, the evaluation team conducted interviews with selected relevant stakeholders where we sought their feedback on the evaluation questions and explored their interests and stakes in the evaluation (the list of stakeholders interviewed is at Annex 8). Substantial insights were provided to assist the development of sub-questions, inform appropriate data collection methods, and provide various other contextual and operational advice.

Many stakeholders sought specific feedback to inform the future direction of the program. Stakeholders placed emphasis on the need to explore sustainability mechanisms and effectiveness of the program from both volunteer and partner perspectives.

This feedback informed the detailed Analytical Framework given below, and the further development of the sub-questions and interview guide.

Limitations and risk mitigation

The evaluation team faced several limitations in carrying out the evaluation. We identified these in advance (as per the Evaluation Plan) and the mitigation strategies used meant these had limited impact on the final deliverables.

Limitation	Adaptive strategies
Limited fieldwork	Travel constraints imposed by COVID-19 meant it was not possible for the Australia-based evaluation team to visit the countries selected for case studies. The inability to reach stakeholders in person may limit the depth and richness of the data collected in some instances. To mitigate this, the evaluation team conducted a thorough desktop review of program-related documentation and used evaluation team members in-country to meet with relevant stakeholders in Solomon Islands, Tanzania, Indonesia, Tonga, and Timor-Leste.
Remote interviews	Remote interviews can be challenging to conduct due to the lack of face-to-face interaction, especially in cross-cultural contexts. Remote/virtual interviews take away the ability for the interviewer to read non-verbal cues, probe further and build the rapport with respondents which is necessary for a deeper conversation. The evaluation team tried as much as possible to build rapport with the respondent, starting with a very clear introduction of themselves and the evaluation, why the respondents had been chosen and seeking oral consent from them to continue. The interviews were informal and semi-structured to ensure the atmosphere and conversation were open and flowing. This allowed the interviewers to probe the respondents and seek further clarity on some of their perspectives. Where possible, in-country Evaluators met face-to-face with stakeholders.
Language barrier	The data collection tools and interviews were designed to be conducted in English. There was thus a potential language barrier especially during interviews with the partner organisations and other in-country stakeholders. In-country Evaluators translated into their local language as necessary for interviews and provided interview notes in English so they could be reviewed by the rest of the evaluation team.
Sample size	The Terms of Reference, and resources available, placed a limit on the number of countries that could be looked at in-depth through the case studies. Given the global scope of the program across 22 countries it would not be possible to cover all countries. The original intent was to sample six countries but to due evaluation team constraints this was reduced to five. While the sample of countries was necessarily limited, the selected countries did offer variability in terms of geographic spread and cultural and policy context. Country case studies were supplemented by data collection at the global level, including from the program's management information system (MIS) which provided data from a much larger, global sample.
Time constraints	The evaluation was pressured for time, with a very large amount of data to collect, code and analyse. A slight delay in finalising the inception report, Christmas holidays in some countries impacting stakeholder availability, a slowed ability to organise some meetings and interviews due to COVID-19 restrictions and remote working requirements, and a requirement to present findings at an early stage all reduced the time for in-depth analysis of the data collected. Despite these constraints, the evaluation team conducted the evaluation as per the Evaluation Plan and deliver the final report to DFAT on time.

Safety and ethical practice

A Safety Plan guided the evaluation (See Evaluation Plan, Annex C) to ensure the safety of the evaluation team and meet the safety standards of DFAT and Tetra Tech.

In keeping with confidentiality and privacy requirements, the Team Leader was responsible for ensuring that data collection and analysis approaches protect participants' privacy by establishing and following credible, ethical evaluation principles. Ethical principles covered the following:

- All participation in interviews will be voluntary, will not create harm to participants during or after the data gathering, and their anonymity and confidentiality will be protected
- Voluntary involvement will be assured by a verbal explanation of the evaluation being conducted and seeking the respondent's written consent to take part in the interview
- The interviewer will inform respondents that they may choose to not respond to certain questions and may end the interview at any time
- We will de-identify data relating to specific individuals
- Individual evaluation participant data will be stored securely and presented in a de-identified manner in all external materials
- Individuals will only be identified in reporting if this is desirable for conveying important findings, and if informed consent is given by the person prior to the written material being circulated
- Interviews and focus groups may be recorded subject to the agreement of participants. This will assist in accurately capturing responses and nuances in perspectives
- Recordings will be used only for this purpose and destroyed immediately after interviews and focus groups have been written up.

The evaluation team conducting interviews summarised the interview discussion points to the respondents and sought their agreement that it was a true reflection of their feedback and reflections.

Approach

This section describes how information was collected and analysed to answer the evaluation questions. It covers the methodological approach underpinning the evaluation, identifies the main types of data and their sources, presents a summary of the analytical framework used to develop findings, and describes how the analysis was done.

1.1   Methodological approach

As set out in detail in the Evaluation Plan, the evaluation aimed to:

- Adopt a **mixed-methods approach** combining different forms of data collection in a phased approach. The evaluation collected both **qualitative and quantitative data** and integrated them into analysis and synthesis. Examples include participation rates and outputs (quantitative), interviews, focus groups and case studies (qualitative). This gives a more comprehensive picture and enables us to triangulate findings and provide robust evidence on the outcomes achieved by the program.

- To extensively review the process and outcomes of the program, the evaluation was anchored in **program theory**, articulated in the form of simple visual and/or written articulations of how different types of activity and volunteering modalities work. The basis of this was the Program Logic from the Program Monitoring, Evaluation and Learning Framework (MELF) 2020.

- Primarily, the evaluation looked at how the program **contributes to outcomes**, rather than seeking to attribute specific outcomes to the program alone. This enables us to deal realistically with complexity, and for DFAT to understand how its efforts fit within a wider perspective.

1.2   Sampling strategy

The evaluation took a stratified purposive approach to selecting stakeholders for interviews and focus group discussions, ensuring that a diverse cross-section of views is represented. The aim was to provide the Evaluation Team with a meaningful overview of the program within the time/resource constraints of the fieldwork. The purposive sampling of interviewees considered logistical constraints and the importance/relevance of stakeholder perspectives to the evaluation questions.

Details as to sampling used for specific tools and processes are provided below.

Key ethical considerations in the sampling related to ensuring a gender responsive approach—ensuring meaningful consultations with all genders and planning these consultations for appropriate times and places. There was also a wider ethical issue to manage—reducing (as much as possible) the tendency for evaluations of this kind to be purely extractive exercises. Where possible, feedback was provided to stakeholders after each meeting.

1.3 Summary analytical framework

The analytical framework describes how data collection and analysis were used to answer the evaluation questions. The analytical framework for this evaluation was guided by and conducted along the Organisation for Economic Co-operation and Development's Development Assistance Committee (DAC) criteria assessing relevance, coherence, effectiveness, efficiency and sustainability of the program, and the evaluation questions associated with each of these criteria.

The table below is a summary of the analytical framework that shows key evaluation questions (aligned to the DAC criteria) and sub-questions under each, assessment criteria and the data sources. The analytical framework is anchored in program theory therefore, testing the theory will be done across the criteria and analysis. The analysis adopted both process and outcome evaluation approaches. By process evaluation, the analysis determined whether the program activities have been implemented as intended (the process) and whether these activities have resulted in certain tangible results (outputs). Through the outcome's evaluation process, the analysis determined the program's effectiveness by exploring whether the program has resulted in the expected outcomes.

This analytical framework was also used as basis of the coding framework where data sources were coded along the criteria and analysed.

Summary of the analytical framework

Question area: relevance of the program design and the extent to which the Volunteer Program has efficiently and effectively made progress towards its End of Program Outcomes

Key:

✓ Yes

✗ No

Components	White and grey literature	Program Documents	Contextual analysis	DFAT Staff perspectives	Government perspectives	Partner organisation perspectives	Volunteer perspectives	Other Partners perspective	Expenditure Data	Results reporting	Quantitative analysis	Map of policy priorities	Program Theory	Case Studies
Assignment level	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓
Country level	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓
Program goal	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓
Program outcomes	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓
DFAT policy priorities	✗	✓	✓	✓	✓	✓	✓	✗	✗	✓	✓	✓	✓	✓
Development outcomes	✗	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓
Capacity outcomes	✗	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓
Soft power	✗	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓
Impact areas (3)	✗	✓	✗	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓
Un / intended	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓
Gender and inclusion	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓
Governance and management	✗	✗	✓	✓	✓	✓	✓	✗	✓	✓	✗	✗	✗	✓
Partnering & influence	✗	✗	✓	✓	✓	✓	✓	✗	✗	✓	✓	✓	✓	✓
Volunteering assignment types	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓

Question area: inform future program directions

Key:

✓ Yes

✗ No

Components	White and grey literature	Program Documents	Contextual analysis	DFAT Staff perspectives	Government perspectives	Partner organisation perspectives	Volunteer perspectives	Other Partners perspective	Expenditure Data	Results reporting	Quantitative analysis	Map of policy priorities	Program Theory	Case Studies
Sustained & resilient	✓	✗	✓	✓	✓	✓	✓	✓	✓	✗	✓	✗	✓	✓
Quality aid	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✗	✓
Adaptability	✓	✓	✓	✓	✓	✓	✓	✗	✗	✗	✗	✓	✗	✓
Modalities	✓	✓	✓	✓	✓	✓	✓	✗	✓	✗	✗	✗	✓	✓
Barriers / enablers	✗	✗	✓	✓	✓	✓	✓	✓	✗	✗	✓	✗	✓	✓
Lessons learned	✗	✗	✓	✓	✓	✓	✓	✓	✗	✗	✓	✗	✗	✓
Improvements	✗	✗	✓	✓	✓	✓	✓	✓	✗	✗	✓	✓	✓	✓

1.4   Data collection

This section describes in more detail the specific methods used for data collection and analysis. These tools were used both with Australia-based stakeholders and those in the five case study countries.

Summary of data collection tools

Method / tool	Description	Outputs
Document review	A desktop review of both operational and reporting data for the program was carried out. The document review process also included a light-touch literature scan to capture learning by other partners about how to support development of the volunteering program, as well as what an effective program of support looks like.	The evaluation team reviewed 220 documents covering a range of topics including the global program strategy, annual and country plans, performance reports, contractual requirements, and financial statements. We used NVivo software to structure the analysis based on the evaluation questions and code data. The document review also fed directly into the comparative analysis (see below).
Program data	The program provided data on output and intermediate outcome level indicators was provided for the period covering January 2018 to December 2020. Program monitoring data is drawn primarily from assignment-level monitoring tools completed by volunteers and partner organisations.	The evaluation team used program data at the country level for the five country case studies as well as informing the analysis in the global evaluation report.
Interviews	The evaluation involved semi-structured interviews with a wide range of stakeholders. The evaluation team conducted interviews following the ethical principles outlined above. Most interviews were conducted online.	The evaluation team interviewed sixty-nine individuals for the evaluation, from several different stakeholder groups: • Volunteers (10) • DFAT staff in Canberra and at post (12) • Staff from the program in Australia and overseas (19) • Partner organisations overseas (18) • Australian-based partner organisations (4) • Staff from overseas partner governments (6)
Focus Group Discussions	Group discussions (with two or more people together) were carried out with a range of stakeholders. Focus groups were conducted online.	A total of 69 individuals particpated in 18 focus group discussions, covering the following stakeholder groups (total number of people in each is provided in brackets): • DFAT staff in Canberra and at post (10) • Partner organisations (17) • Program staff (9) • Volunteers (33) Of the volunteers involved in focus group discussions (FGDs), 27 were in discussions led by the evaluation team in Australia. Participants for these groups self-selected from an email invitation that was sent to all current and former volunteers. Groups were oversubscribed so purposively sampled to maximise variation in age, gender, country of assignment and whether the volunteer was an early return (for reasons within the program's control). Four FGDs were convened with the following groups: Volunteers

Method / tool	Description	Outputs
		who were under 30 at the time of their assignment; who had been on an assignment with a gender equality, disability and social inclusion focus; who had experience of both in-country and remote volunteering; and who had experience both under the current program and a former iteration of it (i.e. Australian Volunteers for International Development). In addition, an FGD was organised with Returned Australian Volunteers Network State Representatives.
		The other six volunteers involved in FGDs were sampled as part of the country case studies (see below).
Volunteer Survey	A short-targeted survey was developed and sent out to all volunteers who had been on an assignment with the program from January 2018 to December 2020.	422 current or former volunteers completed the online survey. The survey went out to all volunteers who had volunteered with the program since 2018 for whom the program had up to date email addresses, were still alive, and hadn't opted out of all communications from the program. This was to 1240 people in total. The survey received 422 responses, or 34% of those who were invited to respond, and 30% of the 1410 volunteers who have been on the program since January 2018. Survey responses were analysed in comparison to the available three-year data set from the program's MIS as that covers all the same volunteer group. The demographic profile of survey respondents was broadly in line with the survey population.

Data collected through these tools was analysed in line with the methodology and analytical framework described above. In addition to informing the main evaluation report, data collected also fed into two additional components of the evaluation.

A **comparative analysis** was produced (see Annex 1), based on the literature scan and interviews with staff from other donor countries. This cross-country analysis was based upon a review of the international literature plus interviews with government staff from volunteer programs in Canada, New Zealand, Norway, and the United Kingdom (the following separate report was also prepared for the evaluation: 'International volunteer programs in Australia, Canada, New Zealand, Norway and United Kingdom: Comparative analysis and future directions'). A targeted component of the literature scan was undertaken to inform a comparative analysis of volunteering modalities, expenditure where possible, emerging issues and responses by other donors and in other contexts together with the identification of international trends and promising practice to be consider for application in future programming.

Five **country case studies** were also produced (see Annexes 2 to 6). Case studies were undertaken of the program's operation in Indonesia, Solomon Islands, Tanzania, Timor-Leste, and Tonga. The purpose of the country case studies was to investigate and compare assignments, capacity, and delivery in countries/regions, to test program theory, and to evidence outcomes. The sample of countries was suggested by DFAT. In each country, the country evaluator carried out several interviews and focus group discussions, with the following stakeholders:

- DFAT Post staff
- Partner Government staff
- Program staff
- Partner organisation representatives (from four partner organisations)

- Volunteers who had been on assignment with some of the selected partner organisations

The partner organisations were sampled on the following basis. The aim of the sampling approach was to capture evidence of 'typical' cases from which analysis of effectiveness can be drawn, while also exploring interesting cases to illustrate different aspects of the program.

The evaluation team used a multi-stage sampling process to select partner organisations in each of the case study countries.

In the first stage, a list of all partner organisations (with related assignments) was taken from the program's MIS (inclusive of all partner organisations supported from start of the program in 2018 to the present). Program managers were then asked to identify seven partner organisations that meet the following selection criteria:

- Identify partner organisations from each of the following categories: 2 X non-governmental organisations (NGOs); 2 X government; 1 X academic/research institution; 1 X international NGO and 1 X private sector = 7. Some variability was required as not all countries had partner organisations in each of these categories
- The selected partner organisations should also meet the following additional criteria:
    - A 'typical' number of assignments since the start of the program in 2018
    - Assignments of a 'typical' duration
    - Working in one or more of the following sectors: government and civil society; education; social infrastructure and services; agriculture and environment; health
    - Partner organisations that have had 'typically' effective assignments (as judged by program manager but verified by monitoring data).

- A criterion for exclusion was of partner organisations that are known to be unavailable or if the partner organisation had been recently sampled for other research (to avoid research fatigue/duplication).

From the shortlist provided by the program for each country, the evaluation team then randomly selected three organisations per country. From the three, two organisations were purposively sampled to capture maximum variation at the global level to ensure a widespread in the 'type' of partner organisations selected, based on:

- Thematic areas (the three impact areas plus health, education, etc.)
- A partner organisation working in a cross-cutting theme relating to gender equality, diversity, and social inclusion
- Types of partner organisation (government; international agency; national/local NGO; etc.)
- A partner organisation with a strategic relationship with an Australian Organisation
- A partner organisation that has had a remote volunteer.

The two partner organisations per country selected through this approach were then the focus of in-depth interviews and the detailed case studies as presented in Annexes 2 to 6. The remaining five organisations that had been identified in the previous stage sampling were all invited to a focus group discussion. FGDs with these organisations were based around the evaluation questions and used to inform the broader stakeholder analysis. In total, 31 partner organisations were involved in interviews and FGDs.

1.5 Data analysis

This evaluation primarily involved rapid qualitative methods of inquiry and used DFAT's own standards coupled with the judgement of the evaluators to interpret stakeholder perspectives. The evaluators made professional judgements based on the evidence gathered and the criteria and standards expressly applied. The evaluation team can substantiate all judgements, i.e. to articulate a clear line of connection from evidence to finding to recommendations.

The analysis was supported by using NVivo, a qualitative data analysis tool. Notes from all interviews, group discussions and documents reviewed were coded in NVivo based on the key evaluation questions. The NVIVO data was then synthesised and key themes identified.

Aide Memoire and Findings workshop. The evaluation team developed an Aide Memoire which outlined preliminary findings and recommendations. These were presented at a Findings workshop on 9 March 2020

which involved over 40 participants, including senior and post DFAT staff, program managers, the Evaluation Advisory Group. This workshop with stakeholders in the analysis phase helped facilitate the sharing and interpretation of the evaluation's initial findings.

Evaluation Team

The core team consisted of four members plus consultants for the case studies. Team members' roles are outlined below.

Table 1: Overview of roles and responsibilities

Name	Role	Responsibilities
Scott Bayley	Team Leader	- Lead team through research, consultation, analysis, and reporting - Lead high level internal stakeholder engagement - Oversee and quality assure research, analysis, synthesis, and reporting - Lead primary and secondary research - Oversee case studies
Susan Hayes	Organisational Development Adviser (Technical Specialist)	- Provide key input to inform stakeholder identification, engagement, and design of consultation strategy - Conduct primary and secondary research - Data analysis and synthesis - Contribute to quality assurance and reporting - Oversee case studies
Jacqui Lord	Research support and DFAT liaison	- Conduct primary and secondary research - Data analysis and synthesis - Support case studies - Key liaison with Posts and other DFAT contacts
Jake Phelan	Research Associate	- Conduct secondary research - Data analysis and synthesis - Support sampling of partner organisations for case studies - Support sampling of volunteers
Devi Miarni Umar **Elias Ponsiano** **Hannah Paleka** **Henrique Soares Ximenes** **Katrina M'Au**	Case studies	- Undertake case studies - Interview partner organisations, volunteers, partner governments and DFAT staff

2.1 Timeline

The evaluation was conducted in line with the phases and timelines set out in the Evaluation Plan.

- The inception phase began in mid-October 2020 and was completed with the delivery of the final Evaluation Plan on 10 December.
- Desktop research and analysis commenced in November, feeding into the final analysis and also the specific deliverable, the comparative analysis report, completed in February.
- Interviews and focus group discussions began in December 2020 and carried on into January 2021.
- The case studies followed by the same timeline, with drafts completed in February 2021.

- The analysis, presentation and reporting phase began in February. An Aide Memoire was submitted to DFAT on the 5th March and an initial findings workshop held on the 9th March. A draft final report was submitted to DFAT on the 28 March.