



Ministry for Foreign Affairs
of Finland MFA



Department of Foreign Affairs of Ireland DFA



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Federal Department of Foreign Affairs FDFA
State Secretariat for Economic Affairs SECO

PEER REVIEW OF FOUR EVALUATION FUNCTIONS

Development Evaluation Unit, Ministry for Foreign Affairs, **Finland**

Evaluation and Audit Unit, Department of Foreign Affairs and Trade, **Ireland**

Evaluation and Corporate Controlling Division, Agency for Development Cooperation, Ministry of
Foreign Affairs, Switzerland

Evaluation Unit, Quality and Resources Section, Economic Cooperation and Development
Division, State Secretariat for Economic Affairs, **Switzerland**

November 2020

Extract for Finland:

Summary

Cross-Agency findings and conclusions

Evaluation in MFA

Consultant:

Patrick G. Grasso, Alexandria/USA

pgrasso45@comcast.net

SUMMARY

Background

Purpose of the Peer Review: Provide an impartial assessment of the evaluation functions of development agencies in Finland, Ireland, and Switzerland (in both SDC and SECO) in terms of the extent to which they meet internationally-accepted standards on the enabling environment for evaluation, the independence of the evaluation system, the credibility of evaluations, and their utility and use.

Methods: Document reviews, key informant interviews, in-depth discussions with evaluation staff, and reviews of selected evaluation reports.

Peer reviewers: Directors of the evaluation units in Finland, Ireland, and Switzerland (SDC) reviewed each other in two-member teams; the director for SECO participated in study design and report reviews. A senior consultant assisted all three teams in common.

Major findings and recommendations

In general, the four units have favorable enabling environments, though with some areas for improvement, as discussed in the respective chapters

Across the four organizations, evaluation is accepted as an important contributor to ensuring projects and programs are meeting their objectives and proving beneficial in terms of development work. However, to the extent evaluation impinges on the related areas of foreign policy and trade there is less understanding of and support for the evaluation function. This is the case in both Finland and Ireland, though not in Switzerland.

Recommendation: To improve the enabling environment for evaluation in non-development areas in the countries where this issue arises, the evaluation units reviewed should consult with similar agencies in other countries, which have been successful achieving this goal.

Across all four organizations independence is generally strong, but some areas could be strengthened, particularly policies on working with outside evaluators and preventing conflicts of interest

The evaluation units all are organizationally independent within their Ministries or Departments, and have demonstrated behavioral independence through candid reporting. However, policies and protocols are somewhat unclear on the extent to which unit staff can participate directly in the conduct of evaluations, and on specifically evaluation-related conflict of interest standards for external evaluators.

Recommendation: The evaluation units should develop policies or protocols that specify the purposes, extent, and limits to active participation by unit staff in the conduct of evaluations (e.g., joining field work, conducting interviews, contributing to products, etc.).

Recommendation: The evaluation units should ensure that their policies or protocols address specific issues of conflicts of interest in evaluation, including those that may arise from external consultants' past or planned future work, and the rotation of staff through the organization.

The credibility of centralized evaluation is high in terms of quality, staffing, other resources, purposes, and use, but decentralized evaluation is less visible, so gets less attention

Stakeholders report that centralized evaluations from the four unit are regard as highly credible. But much evaluation work is decentralized under the control of operating units. There is much less visibility around these evaluations, and even ambiguity about what counts as an evaluation at all.

Recommendation: The evaluation units should work with Management to ensure that evaluation policies are clear on what constitutes a decentralized evaluation and the quality standards and processes for assessing quality are clear.

Finally, at all the organizations evaluation work by the EF was perceived to be useful and was used by operational staff and management, though this was mostly at the policy rather than operational level.

While Senior Management reported that they found the units' evaluation work useful, and actively used it, operational units reported it less relevant to their work. In addition, all four of the units raised concerns about reaching broader audiences for whom their evaluations could be useful.

Recommendations: The evaluation units should strengthen the utility of their evaluations by:

- improving communications and dissemination through:
 - developing a communication strategy that addresses such issues as how to reach internal and external stakeholder audiences;
 - using the planning process to identify opportunities to employ evaluations for multiple purposes, and set priorities on whom the evaluations are expected to influence;
 - including a communication/dissemination plan for individual evaluations from the start.
- strengthening the Management Response process by:
 - clarifying roles and responsibilities for Management and the evaluation unit;
 - setting timelines for both the formal response and agreed actions;
 - enhancing follow-up procedures including tracking of actions taken; and
 - providing for regular reporting on progress and any shortcomings in implementation.

TABLE OF CONTENTS

SUMMARY	i
ABBREVIATIONS AND ACRONYMS	iv
OVERVIEW	1
Introduction: Why Conduct an Evaluation Function Peer Review?	1
Main Issues for the Review	1
Approach and Methods	4
Peer Review Team	5
Organization of this report.....	5
CHAPTER 1: CROSS-AGENCY FINDINGS AND CONCLUSIONS	6
Enabling environment for evaluations	6
Independence of evaluations and evaluation systems	7
Credibility of evaluations	12
Utility of evaluations	13
CHAPTER 2: EVALUATION IN FINLAND	16
Enabling Environment for Evaluation	18
Independence of Evaluations and the Evaluation System.....	19
Credibility of Evaluations.....	21
Utility of Evaluations.....	24
BOXES	
1. Key Roles and Activities to Ensure a High-Quality Evaluation	8
2. Results-Based Management in MFA	14
3. Finland’s Evaluation Help Desk	18
4. Response to Forced Displacement	19
FIGURES	
1. Organization of the Ministry for Foreign Affairs, Finland	16
TABLES	
1. Staffing and Costs for Evaluation Units	11

ABBREVIATIONS AND ACRONYMS

CLP	Core Learning Partnership
DAC	Development Assistance Committee
DFAT	Department of Foreign Affairs and Trade, Ireland
E+C	Evaluation and Controlling Division, SDC, Switzerland
EAU	Evaluation and Audit Unit, Ireland
EF	Evaluation Function
EMS	Evaluation Management Services, Finland
EO	Evaluation Officer, SECO, Switzerland
EU	Evaluation Unit, SECO, Switzerland
EVA	Unit for Development Evaluation, Finland
FCV	Fragility, Conflict, and Violence
IEG	Independent Evaluation Group, World Bank Group
KEPO	Development Policy Steering Group, Finland
MFA	Ministry of Foreign Affairs, Finland
OECD	Organization for Economic Development and Cooperation
SDC	Agency for Development Cooperation, Switzerland
SECO	State Secretariat for Economic Affairs, Switzerland
TOR	Terms of Reference

OVERVIEW

The primary purpose of this Peer Review of the evaluation functions (EFs) of the development agencies of Finland, Ireland, and Switzerland is to provide an impartial assessment of each evaluation function, its strategic fit in the member Ministry, Agency, or Division, and both its good practices and areas for improvement to better support achieving the strategic objectives of development and development evaluation. The review is intended to address four overarching issues, as identified in the Terms of Reference: (1) the enabling environment for evaluations, (2) the independence of evaluation and the evaluation system, (3) the credibility of evaluations, and (4) the utility of evaluations.¹ In addition, each of the EFs also had specific issues that were addressed in its review.

Introduction: Why Conduct an Evaluation Function Peer Review?

The Organization for Economic Co-operation and Development's Development Assistance Committee (OECD-DAC) Network on Development Evaluation (EvalNet) encourages professional Peer Reviews of the evaluation function of its member countries and organizations as a way to "strengthen the evaluation function and promote transparency and accountability in development agencies."² These reviews are carried out to identify good practices and opportunities to further strengthen the evaluation units under review, with a view to contributing ultimately to improved performance in international development cooperation and humanitarian assistance.³

In light of this guidance, the evaluation functions of the development agencies of Finland, Ireland, and Switzerland joined together to conduct this Peer Review. The advantage of combining three Peer Reviews is that it allowed not only findings, conclusions, and recommendations related to each of them, but also cross-agency learning that can be beneficial for the broader development evaluation community. This review is the first Peer Review of the respective EFs outside the regular DAC Country Peer Review process.

Main Issues for the Review

Based on the Normative Framework outlined in Annex II of this report, the following main issues were used to guide the analysis and reporting for the Peer Review. Each chapter is structured around these issues.

Enabling environment for evaluations: The enabling environment includes such factors as requirements for explicit consideration of evaluation findings in developing programs and projects;

¹ These are defined in United Nations Evaluation Group, *Revised UNEG Guidelines for Professional Peer Reviews*, 2016.

² OECD-DAC Network on Development Evaluation, *Evaluating Development Co-operation: Summary of Key Norms and Standards*, 2nd edition, 2010. <http://www.oecd.org/dac/evaluation/dcdndep/41612905.pdf>

³ OECD-DAC Network on Development Evaluation, *Peer Review Reference Guide, 2019-20*, 2019. <https://www.oecd.org/dac/peer-reviews/DAC-peer-review-reference-guide.pdf>.

the degree to which such requirements are carried out in practice; the extent to which operational staff seek input from the evaluation office when considering new programs or projects; and the institutional support provided to the evaluation function by the agency in terms of the adequacy of financial and human resources, access to information, and willingness to cooperate in evaluation activities.

This is an important issue for evaluation because in the absence of a strong enabling environment the independence, credibility, and utility of evaluation work are likely to have little impact on the organization, its work, or its development outcomes, the *raison d'être* of evaluation.

Independence of evaluations and evaluation systems: The credibility of evaluation depends on independence to assure that the findings, conclusions and recommendations are as unbiased as possible. Independence has four major dimensions: organizational, behavioral, avoidance of conflicts of interest, and protection from external influence.⁴

- *Organizational independence* has to do with where in the parent agency the evaluation function is located, especially its reporting relationship to agency Management or some other body, such as an external board.
- *Behavioral independence* refers to the extent to which the unit has been willing and able to carry out its functions with candor, even when its findings may be unwelcome to the broader organization or its Management.
- *Avoiding conflicts of interest* focuses mainly on ensuring that those responsible for conducting and managing any given evaluation are not materially compromised because of current or previous relationships with the activities being evaluated or those in charge of them.
- *Protection from outside influence* involves assurance that the evaluation function is not unduly subjected to pressure from outside, including from Management, governing bodies, or other stakeholders, in the conduct of its evaluations. Of course, stakeholders will have an incentive to try to influence the conduct and findings of evaluations. But the issue here is whether the evaluation function has been given sufficient protection from these pressures to carry out its evaluation work without prejudice.

Credibility of evaluations: The issue of credibility is at heart about the quality of the evaluation products and the system in place to ensure high quality. Work of strong quality is likely to be seen as credible, and without credibility evaluations are not likely to lead to use by the development organization. Questions about credibility focus on whether systems are in place to ensure the professional competence of evaluators and the technical quality of the evaluations (design, methods, and conduct), the transparency and impartiality of the evaluation work, and the clarity of communication of evaluation results.

International development organizations such as the OECD-DAC Evaluation Network (EvalNet), UN Evaluation Group (UNEG), and Evaluation Cooperation Group (ECG) of the international financial institutions (including the World Bank, Regional Development Banks, and International

⁴ These criteria first were derived in World Bank Operations Evaluation Department, "Independence of OED" (February 24, 2003). <http://documents.worldbank.org/curated/en/912291468765325239/pdf/4200ED1Reach.pdf>

Monetary Fund) have developed quality standards for the conduct of evaluations.⁵ These standards are broadly similar and provide a sound basis for assessing the quality of evaluation products. They focus on assuring the competence of evaluators, use of appropriate methods, ethical and balanced treatment of stakeholders in the operations being evaluated, and systems of quality assurance by the evaluation unit.

Utility of evaluations: At the end of the day, evaluation is worthwhile only to the extent that it is useful and used.⁶ The import of this criterion is that to be useful (and used) evaluations must pay careful attention to the:

- multiplicity of audiences they necessarily address, such as parliamentary bodies, governing or oversight boards, senior management, operational staff, development partners, beneficiaries, the research community, and the general public;
- divergent information needs of those audiences, including which should be prioritized; and
- timeliness of their work to ensure it is available when needed, for example to inform emerging strategies or design new operations.

The key issue for the evaluation functions reviewed here is the extent to which their parent organizations have taken on board the findings, lessons, and recommendations from their evaluations and acted on them.

Evaluation needs to be seen as a continuous process, from issue identification through to institutional and individual learning. These criteria help to guide an assessment of how the EFs under review manage that process so as to produce, relevant, independent, credible evaluations that are recognized by the relevant stakeholders as being of high quality and useful – and that are used.

⁵ OECD-DAC Network on Development Evaluation, *Evaluating Development Co-Operation: Summary of Key Norms and Standards*, Second Edition (2010), [link](#); United Nations Evaluation Group, *Norms and Standards for Evaluation* (2016), [link](#); Evaluation Cooperation Group, *Big Book on Evaluation Good Practice Standards* (2012), [link](#).

⁶ Patrick G. Grasso, "What Makes an Evaluation Useful? Reflections from Experience in Large Organizations," *American Journal of Evaluation* 24 (December, 2003). <https://journals.sagepub.com/doi/pdf/10.1177/109821400302400408>. There is an extensive critical literature on this approach, but it has been widely accepted within the development community.

Approach and Methods

In carrying out the review, the following tools and approaches were used:

Document reviews

A key source of information for the review is the documentary record on evaluation organization and management in each institution. Among the documents reviewed were:

- Institutional policies on evaluation conduct and use;
- The Evaluation Function's mandate and related statements on organization, roles and responsibilities, reporting lines, staffing, strategy, and partnership agreements;
- Plans, work programs, and supporting materials;
- Other documents deemed relevant by the Evaluation Function, Management, or other stakeholders.

Key informant interviews

Interviews provide some of the most important information for reviews of this kind because documentary evidence almost never allows for a complete understanding of what is happening on the ground. In particular, stakeholders are far more likely to be candid in interviews than in written comments or other documents. Panel members worked with their respective institutions to identify interviewees for different categories of stakeholders (Management officials, evaluation unit heads, operational staff, external stakeholders). A semi-structured interview guide was developed for each site and interviewee category.

In-depth discussions with evaluation unit staff

In addition to the key informant interviews the Panel members conducted group discussion sessions with staff of the evaluation unit for each organization, including both evaluators and support staff. This was an efficient and effective way to surface internal issues on the operation of the unit and its relationships with other stakeholders that are unlikely to be raised through other forms of data collection.

Review of selected evaluation reports

Finally, the panel reviewed a small sample of reports from each of the agencies for evaluations conducted during the study period; though the terms of reference set this period as 2010-19, the reports reviewed mostly were published after 2015. Key informant interviews helped to identify cases for this review, largely on the basis of their positive or negative receptions or other significant factors.

Peer Review Team

The review was carried out by the heads of four evaluation offices, with teams of two acting as peers for the third; a consultant served on all of the three teams. Members included:

- Patrick Empey, Deputy Director of Evaluation, Evaluation and Audit Unit (EAU), Department of Foreign Affairs (DFA), Ireland.
- Anu Saxén, Director, Development Evaluation Unit (EVA), Ministry for Foreign Affairs (MFA), Finland.
- Christoph Jakob, Deputy Head, Evaluation and Controlling Division (E+C), Swiss Agency for Development Cooperation (SDC), Ministry of Foreign Affairs, Switzerland.
- Johannes Schneider, Head of the Evaluation Unit (EU) and Deputy Head of Quality and Resources Section, Swiss State Secretariat for Economic Affairs (SECO), participated in planning the review and commenting on drafts.
- Patrick G. Grasso, former Advisor to the Director of the Independent Evaluation Department at the World Bank, acted as external consultant.

Peer Review Country Visits

- **Finland (MFA), December 2019**
Peers: Christoph Jakob and Patrick Empey, supported by Patrick G. Grasso
- **Switzerland (SDC and SECO), January 2020**
Peers: Anu Saxén and Patrick Empey, supported by Patrick G. Grasso
- **Ireland (EAU), March 2020**
Peers: Anu Saxén and Christoph Jakob, supported by Patrick G. Grasso

Organization of this report

This report consists of five chapters: a summary of the findings, conclusions, and recommendation from across the four evaluation agencies reviewed, and a separate chapter for each of those agencies. The Annexes provide the Terms of Reference and the Normative Framework for the review.

CHAPTER 1: CROSS-AGENCY FINDINGS AND CONCLUSIONS

One of the reasons for doing this Peer Review across four evaluation functions in three countries was the opportunity for the units to learn lessons and develop ideas for strengthening their own work from directly examining each others' experiences. This chapter presents the major conclusions (in bold) on each of the major issues cited in the Overview that were drawn from this exercise, along with the findings supporting each.

Enabling environment for evaluations

In general, the four units have favorable enabling environments, though with some areas for improvement, as discussed in the respective chapters

Operational managers and staff of the agencies working on development typically are conversant with the need for evaluation, and express support for the work of the evaluation units. Many have served on reference groups that provide inputs to and comments on evaluations at various stages of the work. This has familiarized them with how evaluations are carried out, and appears to have fostered understanding of both the process and the outcomes of evaluation work. Most reported that they have found evaluations useful, though how they have used the results was not always clear.

However, staff with backgrounds in diplomacy and policy, rather than development, often were less clear about, and supportive of, evaluation efforts. These are areas where the routines of evaluation, common to development agencies, are newer and less well-developed. Interviewees did not express hostility to evaluation, but rather a concern about how it might not fit as well with diplomacy and foreign policy activities as with development. A case in Finland highlights the issue; see Chapter 2. The issue also arises in Ireland, though not Switzerland.

However, in some other countries this works well. For example, a peer review of the Policy and Operations Evaluation Department (IOB) in The Netherlands found that its “evaluation system covers policy and operations, process and impact, and a range of interests and needs,” and concluded that it “is an example of a world-class independent evaluation unit.”⁷ Similarly, in Global Affairs Canada’s “whole-of department approach, staff responsible for evaluating international assistance work side by side with colleagues evaluating foreign policy and international trade, and a strategy is in place to enhance learning across branches.”⁸ These cases suggest that in the countries where this issue comes up, the evaluation units reviewed here may be able to learn from such successful cases.

Recommendation: To improve the enabling environment for evaluation in non-development areas in the countries where this issue arises, the evaluation units reviewed should consult with similar agencies in other countries, which have been successful achieving this goal.

⁷ OECD Development Cooperation Peer Reviews: Netherlands 2017, p. 20.

⁸ OECD Development Cooperation Peer Reviews: Canada 2018, p. 2.

Independence of evaluations and evaluation systems

Across all four organizations independence is generally strong, but some areas could be strengthened, particularly policies on working with outside evaluators and preventing conflicts of interest

Organizational independence: The DAC Evaluation Network notes that independence “will best be achieved by separating the evaluation function from the line management responsible for planning and managing development assistance. This could be accomplished by having a central unit responsible for evaluation reporting directly to the minister or the agency head responsible for development assistance, or to a board of directors or governors of the institution.”⁹

The four EFs reviewed in this report provide variants on this guidance, but in each case they report to officials sufficiently removed from the activities being evaluated to meet the standard of organizational independence.

Behavioral independence: The review found that all four units were behaviorally independent. This was confirmed through key informant interviews, discussions with evaluation unit staff, and reviews of evaluation reports. Critical findings and hard-hitting reports were common across all four, and there were no indications that any of them were failing to meet behavioral independence standards.

During the review, the team was asked to consider whether active participation by EF staff in ongoing evaluations would compromise independence, especially behavioral independence. However, independence inheres in the organizations themselves, not in the external evaluators. Participation in the evaluation process does not compromise independence, but can bring to the work institutional knowledge that external evaluators often lack, making the final product more likely to be relevant, useful, and used. As the former World Bank Director General of Evaluation, Robert Picciotto, has argued, “having no connection or shared experience with intended users of evaluation findings, constrains access to information, evinces resistance and inhibits learning.”¹⁰

What is key is that the roles of the internal evaluation managers and external evaluators be clear. For example, Box 1 (below) was developed by the World Bank Group’s Independent Evaluation Group (IEG) as guidance for development evaluation units.¹¹ It offers practical steps that they can implement to manage evaluations without compromising independence. This requires a clear policy and practices that define the purposes, extent, and limits of such participation. However, current policies and protocols across the EFs are not clear on this issue, a gap that should be addressed.

⁹ *Evaluating Development Co-operation*, p. 8.

¹⁰ Robert Picciotto, “Evaluation Independence in Organizations,” *Journal of Multi-Disciplinary Education*, 9:20, pp. 18-31 (2013). https://journals.sfu.ca/jmde/index.php/jmde_1/article/view/373/368

¹¹ World Bank, Independent Evaluation Group, *Managing Evaluations: A How-to Guide for Managers and Commissioners of Evaluation*, 2015, p. 27.

Recommendation: The evaluation units should develop policies or protocols that specify the purposes, extent, and limits to active participation by unit staff in the conduct of evaluations (e.g., joining field work, conducting interviews, contributing to products, etc.).

Box 1: Key Roles and Activities to Ensure a High-Quality Evaluation

The evaluation manager should:

- Ensure that the objectives of the evaluation are clear;
- Maintain ownership of the study by ensuring that decision-making responsibility is retained and that decisions are made in a timely manner;
- Negotiate expectations with stakeholders;
- Monitor the progress of the evaluation and provide relevant and timely feedback and guidance to the evaluator and evaluation team;
- Be open to suggestions from evaluators on possible solutions if problems arise;
- Discuss and ensure agreement on formal and informal communication protocols from the beginning;
- Ensure evaluators have full access to information as early as possible;
- Meet with evaluators and stakeholders to discuss draft reports and revisions;
- Approve the final report and organize a presentation of evaluation findings for stakeholders.

The evaluator (or evaluation team) should be encouraged to:

- Commit to conducting the evaluation within the allotted time frame and budget;
- Provide regular progress reports to the evaluation manager and communicating problems that require the attention of the evaluation manager immediately;
- Discuss the draft report and correct any errors or misinterpretations;
- Respond to comments and finalize the report.

Source: Independent Evaluation Group 2015, cf. footnote 11.

Conflicts of interest: Another area where existing evaluation policies across the four evaluation functions are not always clear is conflicts of interest. Such conflicts may arise from a number of sources:

For example, individuals who may have worked on a project or program years before may seem like good candidates to be involved in an evaluation. However, managers need to be careful because these types of situations can yield biased results. If the manager is not aware of the person's past involvement, the person being approached for the work should be forthcoming and offer to recuse him or herself from conducting or working on the evaluation.¹²

A related form of conflict arises if the individual or firm contracted to conduct an evaluation also has or is likely to be competing for other contracts with the unit responsible for the activity being evaluated. This would raise at least an appearance of conflict, creating risk to the perceived (or

¹² Independent Evaluation Group, *Managing Evaluations*, p. 30.

actual) independence of the evaluation, and raising a reputational risk for the EF. Given the finite supply of qualified development consultants and contractors, these kinds of conflicts are not uncommon. EFs typically rely on institutional procurement policies to fulfill this requirement, as in this reply to an inquiry from a potential contractor:

“If [an organization] is selected as a provider of administrative service for commissioned studies, it needs to follow the laws related to procurement. Considering that the selection of the performer of the commissioned study is finally decided by the [government agency], it is possible for the researchers from the same [organization] to be a performer of the commissioned study. However, the persons functioning as the providers of the administrative service and the persons authorizing their work are themselves disqualified. If the provider of the administrative service proposes experts from the same [organization], the provider of administrative service needs to sign a statement of no conflict of interest.”

However, such broad procurement policies generally do not address specifically the kinds of conflicts that affect evaluations. Therefore, at a minimum it would be useful if EFs supplemented procurement rules with specific policies and protocols for identifying and assessing potential evaluation conflicts, and to mitigate them where possible. These could be worked out with the institutional procurement offices to ensure they are in accordance with institutional rules and applicable statutes. A nice statement of the kinds of issues that should be covered is provided by the US Department of Energy:

Affiliations or activities that could potentially lead to conflicts of interest may include the following:

- Work or known future work for parties that could be affected by the individual's judgements on projects or program developments that the individual has been asked to review;
- Any personal benefit the individual might gain (or benefit of their employer, spouse or dependent child) in a direct or predictable way from the developments of the program/projects they have been asked to review;
- Any previous involvement the individual has had with the program/projects they have been asked to review, such as having participated in a solicitation to the program area that was subsequently not funded, or having a professor, student, or collaborator relationship with the program or its research staff;
- Any financial interest held by the individual (or their employer, spouse, or dependent child) that could be affected by their participation in this review; and
- Any financial relationship the individual has or had with [the organization], such as participation in research grants or cooperative agreements.¹³

Another potential source of conflict is implicit in the personnel rotation system used across the evaluation units. This can make it hard to recruit and retain staff with the right mix of evaluation,

¹³ US Department of Energy, “Program Evaluation: Independence, Conflict of Interest, Openness” (undated).

development, and institutional knowledge to meet the unit's needs, and could provide incentives for staff to be wary of the possible effects of hard-hitting reports on their career prospects.

Evaluation knowledge among staff is important primarily to allow them to contribute to the evaluation work, and to provide strong quality control to assure it is of high technical quality. Development knowledge, gained mostly through experience, helps the team to identify the key issues that require evaluation, and to ensure that evaluation reports are grounded in sector/thematic and country/regional realities. Institutional knowledge is a key to helping the evaluation team conduct its work as smoothly and efficiently as possible, and ensuring that recommendations and lessons from evaluation take account of the institution's capacity to put them into practice. Not everyone in the evaluation unit must have a strong background in all three areas, but the team as a whole needs to cover this ground. And the work environment within the group must foster sharing of relevant knowledge across the team to promote high performance in all three areas. Given the small size of the units, and the institutional rotation policies generally in place, one or two staff with strong evaluation training/experience who can act as a resource for the rest of the staff should be sufficient, but other members should be afforded adequate training to hone their own skills.

This has implications for the recruitment, training, and retention of evaluation staff. The unit may not always be able to recruit many staff with pre-existing evaluation expertise. As a result, it is important that evaluation units invest in training staff to at least a moderate proficiency in evaluation so that they can oversee the work of consultants. The extent to which such training can be provided depends on the resources that can be made available, given the multiple demands on the unit's budget. But this should be given as much attention as is feasible within that constraint. For example, Ireland's EAU staff all have a Masters degree in a relevant discipline, have attended the International Program in Development Evaluation Training (IPDET), and completed other evaluation training courses.

Retention is a more complex issue because without a career path evaluation staff move to other parts of the organization, bringing with them whatever evaluation skills they may have acquired and their institutional knowledge. From the broad view of the organizations, this diffusion of evaluation experience is beneficial, since evaluation work is going on throughout the organization. However, for the evaluation unit this is a serious challenge because it constantly must train new staff. It would be impractical to suggest that evaluation units be exempted from normal organizational assignment policies, or that an evaluation career be established; and in any case such ideas would be beyond the remit of this report. However, in order to promote good workforce planning it may be possible to work with Management to identify one or two positions that can be filled for an extended period by staff with evaluation expertise, including the possibility for promotions. This would serve not only a recruitment and retention need, but also help to re-inforce independence by having the option for some staff to build evaluation careers within the organization.

Recommendations: The evaluation units should ensure that their policies or protocols address specific issues of:

- conflicts of interest in evaluation, including those that may arise from external consultants' past or planned future work; and
- the recruitment and retention of staff with evaluation knowledge and skills, including possible exceptions to general policies of staff rotation in the organization.

Protection from external influence: All of the units reviewed report that, while stakeholders do attempt to influence evaluations through advice and comments, their organizational independence and enabling environments shield them sufficiently to protect against undue influence on the conduct, findings, conclusions, and recommendations from their evaluations.

Part of protection against such influence comes from the adequacy and security of resources available to the evaluation function. Overall, the EF staff report that they are adequately staffed in relation to current workloads, and that budgets are sufficient to carry out their institutional missions. (See Table 1, below.) However, that does not imply that they are working at an optimal level of effort; additional staff and budget resources would allow more thorough evaluation coverage of important issues, if they were made available.

Table 1: Staffing and Costs for Evaluation Units^a

Unit	Staff (FTE)	Budget	Cost per evaluation
Finland	5 full-time staff 8 consultant	€1.5 million	€350-500 thousand
Ireland	6.75	€230 thousand (consultant costs only)	€125-145 thousand (consultants + staff)
Switzerland – SDC ^b	5.4 (all E+C) 2.5 (evaluation)	€660 thousand (consultant costs only)	€70-200 thousand (consultants only)
Switzerland – SECO ^b	1.3	€220 thousand (consultant costs only)	€140-215 thousand (consultants only)

^a The figures in the table are not directly comparable across the units since the costs covered vary among them e.g. Finland's budget is the evaluation unit's entire annual budget; they are presented here only to provide basic descriptive data.

^b Calculated at 1.00 CHF = €0.94.

However, beneath the totals there are human resource issues that create risks to the ability of the units to function effectively. Staff rotation policies of the parent organizations often require experienced evaluation staff to leave for other parts of the organization, to be replaced by new staff with little or no evaluation training or experience. None of the organizations has a career path for professional evaluators, making it difficult to recruit individuals with the relevant skills. In addition, training is mostly spotty for incoming EF staff. All of this leaves the units vulnerable as they oversee evaluation processes and monitor the activities of external evaluators. (see recommendation in the previous sub-section.)

Credibility of evaluations

The credibility of centralized evaluation is high in terms of quality, staffing, other resources, purposes, and use, but decentralized evaluation is less visible, so gets less attention

A review of a small sample of centralized evaluation reports across the four agencies found that they generally met international quality standards. While the quality of individual reports varies, as is to be expected, the officials and others we interviewed expressed a high degree of satisfaction overall with the quality of the work. However, in some cases it was not clear exactly how the units were conducting quality assurance, though clearly there was some routine practice. Units should be sure their processes are well-documented and followed; a good example is Ireland's new quality guidance as discussed in Chapter 3.

However, while institutional evaluation policies typically cover decentralized evaluations, there is much less information on their quality and credibility. For one thing, evaluation policies generally do not define what constitutes a decentralized evaluation, so that a wide range of products, from mid-term reviews, to completion reports, to process reviews, to rigorous outcome and impact evaluations all were described as decentralized evaluations in interviews. The lack of clarity on what constitutes an evaluation is compounded by the lack of visibility of these studies, which makes it difficult to assess their credibility. This can be problematic when findings, conclusions, and recommendations from these reports are used as input into later evaluations, both centralized and decentralized.

The centralized evaluation units reviewed in this report differ in the extent to which they deal with decentralized evaluations. SECO, for example, keeps an inventory of all planned and ongoing evaluations, and publishes all completed evaluations in an annex to its Bi-Annual Performance Report. Finland conducts a meta-evaluation of decentralized evaluations every three years. But none of the four units has direct control over decentralized evaluations, which makes it difficult to maintain quality standards, especially since those overseeing these evaluations usually are operational staff who typically lack expertise in the design and conduct of evaluations.

Some central evaluation units do provide guidance and other technical assistance to operational units that commission evaluations, but this is mostly at the margin, since they do not have direct input to those evaluation processes. In Finland, for example, the evaluation unit helped to facilitate the creation of a help desk through an external management company (see Chapter 2). An assessment of decentralized evaluations was beyond the scope of this review, but this gap in the quality assurance system was readily identifiable.

Recommendation: The evaluation units should work with Management to ensure that evaluation policies are clear on what constitutes a decentralized evaluation, the quality standards for such evaluations, and the processes for assessing quality.

Utility of evaluations

Finally, at all the organizations evaluation work by the EF was perceived to be useful and was used by operational staff and management, though this was mostly at the policy rather than operational level.

Of course, this varies a great deal by individual reports. But the general consensus coming from the interviews is that the work of the EFs is worthwhile and contributes to the conduct of the organizations' work. Many of the operational staff interviewed reported that they perceived centralized evaluations as directed primarily at the concerns of Senior Management, often addressing policy-level decisions rather than operational needs at the project or program level. Country teams, in particular, noted that they often were asked to contribute information for evaluations, but that the final reports did not provide specific feedback or recommendations of use to them, leaving them feeling "taxed" without a commensurate benefit. By contrast, they saw decentralized evaluations as more directly relevant to their work. Implicitly confirming this, senior managers were generally more likely to cite specific uses they have made of the results of centralized evaluations.

This is not a novel finding. In general, it is true that centralized evaluations are geared toward the broader issues facing the organization than any one set of operations. One exception may be country program evaluations, which do address specific country issues, and typically are geared toward informing a new country program under development. Sometimes, this also is true for sector reviews. But in general, centralized evaluation does tend to be directed primarily toward issues confronting Senior Management at most development-related organizations.

This seems a reasonable focus, given that the organizations also carry out decentralized evaluations designed explicitly to address operational issues with projects and programs. In effect, there is a division of labor here that for the most part seems to be able to meet the evaluation information needs of the organizations at multiple levels. Indeed, it is nearly impossible for one report to be equally relevant at the operational and strategic levels, yet sufficiently deep in both areas.

Still, there are some things that the EFs can do to address this issue. For example, when an evaluation includes case studies of specific programs, projects, or other operations, in addition to the main report the EF could issue a set of case studies, with more in-depth information than in the main report. It also could have the evaluators provide briefings or other feedback to the operational teams specifically on their operations that were used in the evaluation. Some of this already is being done, but it could be more systematic.

One issue of common to all evaluation units is how to improve the communication about and dissemination of evaluation findings, conclusions, lessons, and recommendations to a range of audiences that might have an interest. In the subsequent chapters there are specific discussions of this issue for each unit. But the more general finding is that all evaluation units should recognize the need for a regularize set of communication activities to disseminate their work to audiences likely to be interested in and able to put to use their work. This requires both a general

communication policy and procedures to carry it out, but also a communication plan for each evaluation that is developed early in the planning process.

An important facet of use is the Management Response. This is the organization's formal process for taking on board the results of evaluation, accepting accountability, and demonstrating a capacity for learning. Management acceptance of recommendations and implementation of actions responsive to them is the most obvious example of utility. (See Box 2, below, for an example.) During the review, three issues about the Management Response were raised to one degree or another across the organizations: (1) lack of clarity about who is drafting and delivering the response; (2) frequent delays in delivering the response; and (3) difficulties in tracking implementation of agreed actions included in the Response.

These issues partly reflect process ambiguities that could be addressed through operational protocols. But they likely also are symptoms of a deeper problem with Management's ability or willingness to accept and implement evaluation recommendations.

The World Bank's IEG, confronting similar problems, conducted a series of case studies several years ago to determine what led to actual implementation of recommendations. They found that several factors are crucial:

- in-depth discussion with management during the drafting of recommendations;
- credibility of the evaluation results;
- sense of shared ownership over the evaluation and the findings;
- quality of the recommendations, specifically in terms of their timeliness, how actionable they are, and their cost-effectiveness, clarity, and coherence;
- advocates/champions supporting the adoption of the recommendations; and
- institutional incentives and accountability for adopting the recommendations.

Box 2: Results-Based Management in Finland's MFA

The Development Evaluation Unit produced three reports on Results-Based Management (RBM) in MFA's development work over the past decade. Two early studies, *Results-Based Approach in Finnish Development Cooperation* (2011) and *Evaluation of Finland's Development Policy Programs from a Results-Based Management Point of View 2003-2013* (2105) focused on helping to guide the development of a robust RBM system in MFA.

In interviews, a number of senior officials cited these as "high-quality" reports and good examples of how evaluation can influence MFA's work. One noted that the 2011 report, in particular, "provided a wake-up call to Management." Even though not all the recommendations were implemented, the report spurred constructive discussions across MFA that led to subsequent Management actions.

The most recent report, *How do we Learn, Manage and Make Decisions in Finland's Development Policy and Cooperation* (2019), looked at how much progress had been made, and found that RBM was having a "moderate level of influence on results information on both learning and decision-making", but also cited a number of continuing weaknesses, including a "lack of institutionalization of evidence-informed learning and decision-making at MFA."

Based on these findings, IEG developed a new process for developing recommendations, set a time limit of 90 days following discussion of the report at the Committee on Development Effectiveness for Management to submit detailed action plans and timelines for implementing recommendations accepted as part of the Response, and began reporting to the Committee on a quarterly basis the extent of implementation by Management.¹⁴ The World Bank's tracking system can be more resource-intensive. However, SECO has deployed a more streamlined approach that is discussed in Chapter 5 (see Table 4 in that chapter).

Recommendations: The evaluation units should consider strengthening the utility of their evaluations by, *inter alia*:

- using the planning process to:
 - ensure that the objectives and scope of the evaluation are clearly aligned with the information needs of identified internal or external stakeholders and audiences;
 - set priorities among those information needs to ensure maximum utility (see, for example, Table 3 in Chapter 3);
 - work with the evaluation team to clarify what data are needed to meet these information needs;
 - agree with the evaluators on how EVA staff can participate in evaluation activities so they can ensure the work remains focused on those information needs; and
 - identify opportunities to employ evaluation findings and products for multiple purposes (including inputs to other evaluations, where appropriate) involving multiple stakeholders and audiences.
- improving communications and dissemination through:
 - developing a communication strategy that addresses such issues as how to reach internal and external stakeholder audiences;
 - using the planning process to identify opportunities to employ evaluations for multiple purposes, and set priorities on whom the evaluations are expected to influence;
 - including a communication/dissemination plan for individual evaluations from the start;
- strengthening the Management Response process by:
 - clarifying roles and responsibilities for Management and the evaluation unit;
 - setting timelines for both the formal response and agreed actions;
 - enhancing follow-up procedures including tracking of actions taken; and
 - providing for regular reporting on progress and any shortcomings in implementation.

¹⁴ World Bank, Independent Evaluation Group, Managing Evaluations: A How-To Guide for Managers and Commissioners of Evaluation, 2015, pp. 41-47.

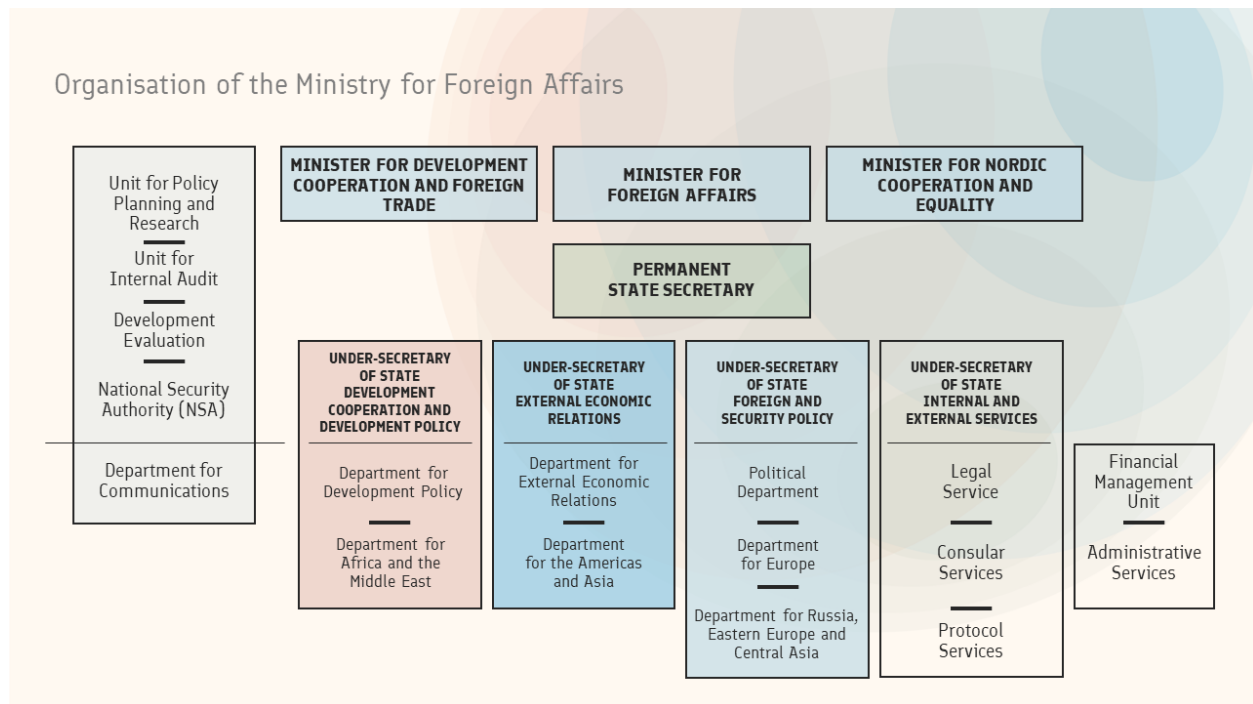
https://ieg.worldbankgroup.org/sites/default/files/Data/reports/ecd_man_evals.pdf

CHAPTER 2: EVALUATION IN FINLAND

The Ministry for Foreign Affairs (MFA) manages and co-ordinates most of the Finnish development co-operation programs. Under the MFA, the Department for Development Policy is responsible for providing overall guidance on the implementation, planning, and monitoring of Finland's development co-operation policy, and holds direct responsibility for the operational activities directed to multilateral and civil society organizations, development co-operation, and humanitarian aid. Regional departments are responsible for the implementation of bilateral co-operation.

The institutional arrangements and responsibilities for evaluation are defined in the Decree on the Ministry for Foreign Affairs 550/2008(1280/2013) and the Evaluation Norm 1/2015. The Unit for Development Evaluation (EVA) is responsible for the development of the evaluation system, commissioning policy and strategic evaluations, and ensuring their effective use. It became an independent administrative unit in January 2014, reporting to the Under-Secretary of State for Development Cooperation and Development Policy. (See Figure 1.)

Figure 1: Organization of the Ministry for Foreign Affairs, Finland



Through its participation on the Intervention/Project Quality Assurance Board of the MFA, EVA is expected to ensure that evaluations are used for intervention planning, that the evaluability of new interventions is high, and that an initial evaluation plan is integrated into the funding proposal. Furthermore, EVA also is part of the Development Policy Steering Group (KEPO) of the MFA, and an expert member of the national Development Policy Committee to facilitate participatory evaluation planning and provide advice on evaluation-related issues.

The planning cycle covers three years, the first of which is elaborated in detail. The ministry-level evaluation plan is prepared by EVA in co-operation with policy makers, senior management, as well as implementing units. The evaluation plan is discussed with Development Steering Group (KEPO) and later presented to the Under Secretary of State for approval. Both centralized and decentralized evaluations are included in the plan. However, the actual planning and implementation of the decentralized evaluations is done by the implementing units themselves.

Uniquely among the EFs included in this review, EVA outsources conduct of its evaluations to a company that implements all stages of the evaluation process, from terms of reference (TORs) to final reports, through an Evaluation Management Services (EMS) contractor, based on a concept note drafted by EVA.

The purpose of this arrangement is to improve the quality of evaluations by securing the best possible evaluation professionals without bearing the time and resource burdens of the tendering process, thus maximizing the flexibility, efficiency and effectiveness of the MFA in planning and commissioning evaluations. The EMS contract is renewed each four years through competitive bidding.

EVA staff and other MFA interviewees report a high degree of satisfaction with both the quality and management of evaluations through this system. One especially appealing aspect is that the contractor is responsible for identifying team leaders and other members for each evaluation, using their networking capacity. Final decisions rest with EVA, but off-loading the search process is an especially useful element of this contract process.

One issue with this arrangement is that it may create some ambiguity about the role of the internal evaluation manager (EM) vis-à-vis the team and team leader. The EM's role is not clearly specified in the framework contract, so a wide range of possibilities could be compatible with it. A minimalist position would be that the EM plays essentially an administrative role, overseeing necessary paperwork, ensuring contract compliance, and the like. A more active role could include joining the team on field missions, primarily as an observer, to ensure the quality of the work being done.

Experience among consulting firms and other organizations involved in this kind of outsourcing indicates that Unit participation in the evaluation process as part of the managerial function—such as observing field missions, interviews, team meetings, and the like—can be beneficial in helping to ensure the quality of the finished product, while allowing the EM to identify any emerging problems early enough to rectify them. This intermediate role requires little background in technical evaluation skills, but at least some management experience likely would be helpful, particularly in helping to control costs, though staff often lack such skills.

Recommendation: EVA should ensure that its EMs either have experience in contract management consultants, or provide training in such skills for those who need it.

Enabling Environment for Evaluation

The enabling environment includes such factors as requirements for explicit consideration of evaluation findings in developing programs and projects; the degree to which such requirements are carried out in practice; the extent to which operational staff seek input from the evaluation office when considering new programs or projects; and the institutional support provided to the evaluation function by the agency in terms of the adequacy of financial and human resources, access to information, and willingness to cooperate in evaluation activities.

Box 3: Finland's Evaluation Help Desk

The Evaluation Unit in Finland's Ministry for Foreign Affairs has a contract with an external consulting firm for evaluation management services. In addition to managing the independent evaluation process for the Unit, it also supports other operational units in their decentralized program evaluation activities by providing capacity development tools and activities, evaluation guidelines, and a help desk services that provide particularized assistance. This function has proven popular among operational units. Which have provided positive feedback on its utility.

Overall, there is a strong enabling environment for development evaluation within MFA, though with areas that could be improved. This is conveyed, for example, by an Evaluation Policy requirement that every development project be evaluated at some point within its lifetime. Such evaluations may come at different stages, frequently as mid-term or completion reviews (Project Final Evaluations). In a few instances, impact evaluations have been commissioned. These evaluations are planned and implemented through the units managing the projects, not EVA. This has the benefit of making operational staff comfortable with the idea of regular evaluation, and contributing to their understanding of the need for and usefulness of evaluation. However, EVA does provide some support for these evaluations, including training and help desk services. (see Box 3.)

Helping to strengthen this environment are some tools that EVA has prepared to provide guidance applicable to both centralized and decentralized evaluations. Particularly noteworthy is a web-based *Evaluation Manual* that combines documents, graphics, and videos to walk users through a wide range of issues, from an explanation of what evaluation is and why it is important, through discussions of evaluation standards and ethics, evaluation questions, methods and designs, data collection and analysis, managing the evaluation process, to reporting and follow-up, and much else. It is both user-friendly and encyclopedic, and can serve not only as a training tool but as a ready reference for those planning and conducting evaluations.¹⁵

The Peer Review team found a notable culture of evaluation within the organization. Interviews with senior managers, staff responsible for managing projects and programs, and other officers found a broad understanding of and commitment to evaluation as a tool for conducting and improving MFA's development work, including the centralized evaluations conducted by EVA. Importantly, this extends to budget support for EVA work, all the way to the level of the Parliament, as discussed below. Finally, the size of the Ministry is small enough that staff throughout are

¹⁵ EVA, *Evaluation Manual*. <https://um.fi/development-cooperation-evaluation-manual>.

familiar with EVA's evaluation work through informal interactions, as well as more formal communications. This can be a double-edged sword, however, as discussed below.

One issue that has arisen is the scope of EVA's evaluation mandate. The Finnish International Development Agency was merged into the MFA several years ago, in part because of the implicit relationship between official development assistance (ODA) and broader foreign policy and trade-related activities. However, this necessarily means that in some cases evaluations of ODA work will touch on other these other areas of MFA responsibilities. In general, evaluation is not routine in those areas, and this can raise issues about the appropriate scope of evaluation work.

For example, a recent evaluation on forced displacement provided insights into Finland's work in this area, and the nexus between displacement and humanitarian aid. The findings and recommendations from this evaluation necessarily go beyond development assistance work and touch on issues of foreign policy and relationships with international organizations in non-development domains. (see Box 4.)

This is important because as international development programs increasingly intersect with those domains, evaluators will find themselves navigating what one interviewee called the "tricky" terrain that defines both the boundaries of their competence and their ability to formulate recommendations that can be implemented by all the relevant agencies. In practice, however, there is no need to change the formal mandate for EVA, since evaluations cover all departments within MFA already.

Box 4: Response to Forced Displacement

The *Evaluation on Forced Displacement and Finnish Development Policy* (2019), was designed to "assess how coherently Finland's development policy and its targets relating to forced displacement have been implemented and how the coherence could be enhanced." A number of interviewees cited this report as showing how the interaction of development and foreign policy activities can affect the work of evaluation units.

The report found that Finland, though "a highly respected development and humanitarian actor and advocate...has not developed clearly formulated and well-established approaches that effectively inform its policy making and programs for development cooperation and humanitarian assistance in a coherent and comprehensive fashion" in addressing issues of forced displacement and the humanitarian-development nexus.

Recommendation: EVA should consult with other evaluation units, such as those in The Netherlands and Canada discussed in Chapter 1, that have successful dealt with the non-

Independence of Evaluations and the Evaluation System

Organizational independence

As noted above, EVA reports to the Under-Secretary of State for Development Cooperation and Development Policy, who does not have responsibility for operations undertaken by the regional departments. This arrangement meets the criteria for organizational independence, though it might be stronger if the unit reported to the State Secretary or Minister for Development Cooperation and Foreign Trade.

Moreover, the budget allocated to the Unit appears to be adequate to meet its mandate. In interviews, EVA staff reported that they did not feel constrained by the budget to compromise the breadth or quality of the evaluations on which they were engaged. Interviews with interlocutors between MFA and the Parliament noted the strong support for providing adequate budgets for evaluation at all levels of government. This level of support also reinforces behavioral independence by removing a potential source of pressure on the unit

Behavioral independence

A review of a sample of reports and responses from those interviewed consistently emphasized that the Unit exercises its function with a high degree of independence. Indeed, the case of the evaluation of forced displacement discussed above is a good example of the assertion of independence even when questions were raised about whether EVA was going beyond its mandate. No document the Peer Review team examined, and no interviewee, cited any instance of independence being compromised behaviorally.

EVA annually develops a comprehensive three-year evaluation plan in consultation with potential users of evaluation; the plan includes decentralized evaluations to be undertaken by operational units. This plan, and any subsequent changes, is presented to the Development Steering Group (KEPO) for discussion, but approved by the Under-Secretary of State, who is responsible for ensuring the independence of the unit. The comprehensive plan for centralized evaluations is binding for the first year and indicative for the following two years. General evaluation principles guide the plans, including that the results will be available when the information is needed, though in practice this objective is often not met because of the lengthy time for evaluations to be completed. The plan also discusses related training.

In addition, a Development Policy Committee helps to bolster the Unit's independence by bringing together institutional members from academia, stakeholder groups, Ministries and Parliament to discuss evaluation reports and "translate" them into a more user-friendly form for policy-makers. The Committee also helps to gain support for EVA's budget, and can look into past recommendations to see what has been done in response. There may be scope for the Unit to make use of this Committee strategically to further promote its messages and encourage action on its recommendations, as discussed below.

Avoiding conflicts of interest and outside influence

Apparently, there is considerable interest in EVA's work among the extensive network of non-governmental organizations in Finland active on development-related issues. They do express their views, but there is no evidence that this has had any compromising effect on the independence of the evaluation process or products. Internally, interested units also may wish to influence EVA evaluation findings and recommendations, not only through the legitimate formal channels, such as commenting on draft reports, but by taking advantage of the relatively small size of the organization to exert informal influence. Despite this, the review did not surface any evidence of loss of independence from these possible sources.

However, there are personnel issues associated with rotation of staff from across MFA into EVA, as with other assignments. In addition to the potential for conflicts of interest from both past work and future opportunities this implies, these staff usually do not come with training in or experience with evaluation. This is discussed in Chapter 1, which includes recommendations on policies regarding conflicts of interest, as well as staff recruitment and retention.

Credibility of Evaluations

In interviews with operational managers and staff, the Peer Review Team was told that the framework agreement had helped improve the quality of EVA evaluations even as it has allowed the quantity to increase. Concomitantly, EVA staff reported that their administrative and procurement workload had gone down, leaving more time to attend to substantive tasks related to the evaluations. Both inside EVA and across MFA the framework agreement has been received positively.

However, this arrangement carries a number of risks for EVA. First, the EMS coordinator provides EVA with a list of three Team Leader candidates with a proposed approach from each. While EVA makes the final selection of Team Leader it is somewhat dependent on the external EMS coordinator for identifying a short list, raising the risk of real or perceived conflicts of interest, as discussed above.

Second, it also leaves EMS with the lead role in defining the scope of work, although EVA does produce a Concept Note at the start of the evaluation, and must approve the Terms of Reference and scope. This has important implications for the cost, timeliness, and utility of the evaluations undertaken. EVA's costs per evaluation are considerably higher than those for the other units. The reports also tend to be longer, denser, and more complex, sometimes appearing more as research products than evaluations *per se*. This likely reflects the fact that the work typically is led by university faculty members. (As discussed below, this also has implications for the utility and actual use of evaluations.)

A major way EVA could address this issue is by taking a more active role in the design of evaluations. The Concept Note is a vehicle that could be especially useful in directing the external consultants as they prepare the Terms of Reference and scope of work, proactively stemming the "mission creep" that has made some of the evaluations more expansive and expensive than necessary. It therefore should provide specific guidance on:

- the specific purpose and objectives of the evaluation, including a clear statement of why these issues are important to the work of MFA, not just to the field in general;
- the evaluation questions to be addressed, and any limitations on the scope of work to ensure that the focus remains on those questions;
- the expected use of the evaluation findings, conclusions, and recommendations, including, where possible, identification of specific users (such as MFA units or teams, external partners, policy-makers, and so on); and

- preliminary plans for how the evaluation results will be communicated and products disseminated.

The Concept Note thus can be a strong mechanism for guiding the work of the evaluators, including decisions on the Terms of Reference and scope of work. The TORs normally should conform to the Concept Note rather than adding to or subtracting to it, unless the evaluators make a compelling case that changes are needed to meet the stated purposes and objectives of the evaluation. Any agreement to such changes should be made only by the EVA Director, based on careful consideration of the trade-offs in terms of timeliness, costs, and staffing.

There may be some reluctance among EVA staff to do this because (1) staff without evaluation experience may be unwilling to challenge outside evaluators on technical issues, although EVA has made use of “critical friends” from outside the unit to advise the Evaluation Manager on these issues; and (2) there is some concern that a more active role would compromise the independence of the evaluation. But these concerns should not be determinative.

For one thing, while external consultant may have subject-matter and methodological expertise, EVA has a higher level of institutional knowledge. This means EVA readily can find out from internal MFA units with stakeholder interests in any evaluation what kind of information they will find most useful. A good place to start would be with the reference groups typically set up to advise on each evaluation. Reference groups are made up of MFA staff, and have input at five points:

- during preparation of the Concept Note;
- at the start-up meeting with the evaluation team;
- when commenting on the draft Inception Report;
- as part of a validation exercise when the team presents initial findings/data, and draft recommendations;
- in commenting on the draft final report.

There has been little research into the structure and use of reference groups, even though they are widely employed in evaluation work, including by the units covered in this review. So it is not possible to characterize any “norms” for such groups in terms of size, composition, skills, or roles and activities. However, several years ago the American Evaluation Association did publish a volume in its *New Directions in Evaluation* series that included some case studies and a proposed model for these kinds of groups.¹⁶ The authors of the latter note that, “who is invited to participate in [a reference group] depends on the needs of an evaluation study and the context within which the study is taking place.”¹⁷ They identify three main types of participants:

¹⁶ R. VeLure Roholt and M. L. Baizerman, “A Model for Evaluation Advisory Groups: Ethos, Professional Craft Knowledge, and skills,” in R. VeLure Roholt and M. L. Baizerman (eds.), *Evaluation Advisory Groups. New Directions in Evaluation* 136, pp. 119-27 (2012).

¹⁷ *Ibid.*, p. 120.

- experts with technical evaluation expertise (e.g., design, data collection and analysis);
- stakeholders with interest in the outcome of the evaluation;
- individuals who can help craft and disseminate the evaluation messages.

In the context of MFA and the other organizations reviewed here, technical evaluation expertise tends to be in short supply internally, so other resources may need to be brought to bear. EVA has introduced “critical friends” with evaluation and/or substantive expertise to assist the EMs in designing and managing evaluations. The stakeholders of interest are most likely to be MFA staff of the office(s) whose work is being evaluated, others with a direct role in the country(ies) or substantive area(s) covered by the evaluation, and perhaps those with responsibility for institutional policies that might be involved. Evaluation units typically do not have staff with singular responsibility for promoting their messages, but the parent organizations do, and these can prove helpful, even if not included on a reference group.

Properly constructed and used, therefore, reference groups can supply useful inputs to the design, implementation, and final product of an evaluation; give entrée to knowledgeable individuals and groups with a stake in the issues being evaluated who otherwise might not be as available to the evaluators; and engage stakeholders in the evaluation process to both foster the credibility of the results and create opportunities to disseminate them to the broader relevant community(ies) of practice. In addition, operational staff interviewed for this review who had participated in reference groups reported that they found the experience generally useful, both because they learned about issues relevant to their work, but importantly because they developed an appreciation for the evaluation process itself, giving them a clearer understanding of how it could contribute both to their own work and to the broader efforts of MFA.

Necessarily, the size and composition of such a group will vary depending on the scope and complexity of the evaluation. A review of work in one sector in a single country might require no more than three members, while a broad evaluation of a multi-sector effort across numerous countries would require many more. There is no simple rule of thumb here, except to recognize that there usually is a tradeoff in the size and efficiency of any group. Limiting stakeholder membership to the key issue areas, for example, might be helpful in managing this tradeoff, but in the end this is a matter of organizational politics as much as anything else, so knowledge of the organization and the key points of possible resistance/support is an important consideration.

Beyond that, a more active role for EVA staff in the conduct of evaluations, including participation in field visits or other evaluative work, does not undermine independence, and can serve both to better manage the evaluation and to help staff hone evaluation skills. Independence inheres in the evaluation unit itself, whether evaluations are conducted externally (as in the unit under review) or internally (as at the World Bank). The key is to define the roles of internal and external staff clearly so that expectations are harmonized and conflict minimized, and to have a strong commitment to the unit’s independence by the organization as a whole.

Recommendation: EVA should develop specific guidelines for:

- the selection of members of reference groups to ensure that a skill mix reflecting evaluation design and methods, stakeholders with substantive knowledge of the subject matter and countries covered by the evaluation and, if possible, with skill in shaping and disseminating messages; and
- conduct of the reference group, including which tasks are expected to involve formal meetings and which may not;
- clarifying the role EVA staff may play in working with evaluators in conducting field work and other evaluative activities.

Utility of Evaluations

Evaluation ultimately is expected to be used to achieve practical results, such as improving policies, programs, projects, and other activities in light of findings, conclusions, and lessons. In MFA, as in other organizations supporting international development, the key ways in which this is expected to happen are institutional learning and implementation of recommendations. Thus, it is important that the evaluation reports are not only independent and credible, but that they address issues important to MFA decision-makers in a timely way and with clear messages. This review found some issues in this regard, however. At EVA, some of the utility issues center on how managing the scope and complexity—and thus cost—of evaluations to improve utility may affect the independence of evaluations, how to improve utility through the management response process, and ways to communicate/disseminate learning from evaluation to audiences outside MFA.

Scope, complexity, cost, and independence

EVA's evaluations are more expensive than those at the other organizations. Interviews with MFA managers and staff, as well as a review of a sample of reports, suggest that in part these costs reflect how the evaluations are conducted. Compared with the other organizations, EVA evaluations tend to be more ambitious in scope, more complex in methods, and longer in duration. All of these drive costs, and also have implications for the utility of the resulting products.

In reviewing sample reports, one of the things that stands out is the amount of space in EVA reports taken up by discussions of the underlying issues and the methods employed. For example, the report on a 2018 meta-evaluation on the treatment of women's and girls' rights in Finnish development policy included a nine-page discussion of how the meta-evaluation was conducted, and ten pages on the context for the evaluation, in addition to several pages on the underlying theory of change.¹⁸ This kind of detailed exposition would be of interest to evaluators, researchers, and some readers interested in the policy background. But both sections clearly could have been much shorter, with the more detailed information moved to an appendix or made available online

¹⁸ Evaluation on Improvement of Women's And Girls' Rights in Finland's Development Policy and Cooperation, 2018. [https://www.shareweb.ch/group/Peer-Review-Evaluation-Function-between-IR-SF-CH/Shared%20Documents/Documents%20Finland/Central%20Evaluations/Evaluation Women and Girls rights NE TI 2018.pdf](https://www.shareweb.ch/group/Peer-Review-Evaluation-Function-between-IR-SF-CH/Shared%20Documents/Documents%20Finland/Central%20Evaluations/Evaluation%20Women%20and%20Girls%20rights%20NE%20TI%202018.pdf)

for interested readers. By front-loading these long discussions the report created a barrier for readers primarily interested in the analytical findings. Moreover, in interviews operational managers often characterized EVA reports as “theoretical”, almost certainly reflecting the heavy emphasis on the theory of change, broad context, and evaluation methods front-loaded in the reports.

What this suggests is some lack of clarity on the part of the evaluators about who was the primary audience for the report: MFA managers. It is not clear from the evidence, but some interviewee comments suggest that the focus on methods in the reports reflects the time and attention the evaluators spent on refining the methods. This possible misdirection of attention is reflected as well by the relatively long time-frames for some EVA evaluations, which often means that by the time they are completed interest that might have been high at the start might have waned or key decisions might already have been made.

At the same time, the high cost of many EVA evaluations limits the ability to conduct other work, and therefore the evaluation coverage for MFA. EVA’s dependence on the evaluation team leader for designing evaluations contributes to this problem, and the lack expertise among EVA staff, also appears to contribute to the high costs of the work.

A conclusion here is that EVA could exercise more control over the evaluations it commissions through the EMS. It likely can commission some evaluations that are methodologically simpler and less costly. But that would require that the ambitions of the work be more modest, as well. Evaluating big, complex programs is inherently more difficult and more expensive than doing reviews looking at management of projects, for example. From interviews it does not appear that there has been a discussion between EVA and MFA management on whether a greater diversity of evaluations along these lines would be useful, but that would be a logical way to begin addressing the issue.

EVA also has raised concerns about whether focusing too much on the utility of evaluations is a risk to evaluator independence. And, in fact there have been debates about this issue within the evaluation community for decades. But in general development evaluators across the international financial institutions, United Nations agencies, and OECD-DAC long have endorsed utilization as the key outcome of evaluation because their mandates are to use evaluation to promote accountability and learning, not to conduct basic research. The key is to conduct evaluations that address important issues for the parent organization, in a timely way, with high quality and credibility. Decades of experience demonstrates that this can be done without sacrificing independence. Ultimately, control over the scope, complexity, and cost by EVA are compatible with independence because, as noted above, independence inheres in the unit itself, not in the ability of external contractors and consultants to exercise unfettered control over the evaluation process.

Recommendation: EVA should explore ways to diversify its portfolio of evaluation products so as to address important issues more efficiently through:

- conducting some less complex and expensive evaluations;
- establishing through its EMS agreement requirements that evaluators employ the least-cost methodology that is adequate to answer the evaluation questions; and
- redesigning reports so that much of the technical material is moved to annexes or appendixes, allowing readers to focus on findings, conclusions and recommendations.

There are many examples among the development agencies, but in the end for the process to work it must be particularized to MFA, with buy-in by both EVA and Management.

Broader communication and dissemination

Finally, EVA is interested in extending the reach of its communication and dissemination activities to a broader audience in order to influence development policy through public discourse and to raise the visibility of MFA's development work, including achievements. Historically, this has been a somewhat neglected area among official evaluation units, partly because of reticence to participate directly in public debates, which could be seen as "taking sides" and impinge on the perception of objectivity and independence, and partly because most evaluation units lack the human resource skills to manage relations with the media and other actors of interest.

The main way to get media coverage of evaluation reports is to establish ongoing relationships with reporters and editors interested in the subjects of the evaluations. For example, when IEG decided to take advantage of a new, more outward-looking disclosure policy it identified a reporter at the *Financial Times* who had written a number of articles about World Bank work. They invited him to speak at a department retreat to explain to staff what makes a report of interest to reporters, and then kept in touch with him, alerting him to evaluations that might be of interest. This proved successful, and soon, working with a media relations professional, they had established a number of such relationships.

Identifying appropriate journalists could be done through a number of channels. Creating a contact information database of any who write about an EVA report or about Finnish development assistance, for example, which could be used to send notifications of new reports or short notes on important findings, using email or social media (reporters tend to be heavy users of Twitter, for example). Personal meetings with journalists are an important way both to establish relationships and to learn what specific issues are of interest to them. To do this kind of work, however, normally requires someone with experience in this area of work. That does not mean a full-time staff member would be required; given the volume of EVA output a part-time consultant likely would be sufficient.

But before pursuing such an effort, EVA would need assess the possible trade-offs. First, MFA has a Department of Communications as well as a communication policy and strategy. It would be necessary to determine whether those would require that such communications go through this

Department. If so, there is a risk that EVA's independence could be compromised, especially if that Department would control the content of any announcements, news releases, or other communications, or be responsible for screening contacts with journalists. Second, this work would require at least some financial resources, as well as the time and attention of the EVA Director. That's an opportunity cost in relation to the rest of the unit's work that would have to be balanced against the potential gains in terms of influencing policy and program decisions and informing stakeholders and the broader public. Third, EVA would have to weigh the reputational risk that it would draw negative attention to its work or that of MFA, however unintentionally.

Another outlet available to EVA is the Development Policy Committee, a group of about 80 individuals representing academe, stakeholders, and ministries. The Committee discusses EVA reports, "translates" them into non-technical terms for discussions with parliamentarians and other policy-makers, and does some checking on recommendation follow-up. It is possible to build on the Committee's membership and existing network of contacts to amplify EVA messages. To do this strategically, EVA would have to develop a clear statement of its communication/dissemination goals, identify the kinds of audiences it most needs to reach to meet those goals, and assess the extent to which the Committee can act as an appropriate conduit. The Committee could be a valuable resource if it is engaged with a strategic vision.

These suggestions should not be taken as formal recommendations, since a good deal of additional work would need to be done to assess the desirability and feasibility of carrying out such activities. Rather, these should be seen as suggestions on how to proceed, if EVA decides to explore these directions. However, EVA should already be considering ways to serve its multiple audiences.

All of the above can be summarized as advice to EVA to consciously consider all the ways its evaluation work can be used beyond the issuing of specific reports. For example, case studies conducted as part of an evaluation of a country-level program could also provide useful input to sector studies, and also could be published as stand-alone products that would serve the needs of both country and sector teams, as well as partner country officials, NGOs, academicians, and journalists with an interest in the specific subject. To do this effectively, however, EVA needs to assess the various audiences for any given evaluation and determine which are most likely to actually use the information, what format would be most useful for that audience, and even when the information would be most useful. This should be under consideration from the start of any evaluation and reconsidered over time routinely as conditions may change. A modest investment of time on such activities can yield a big return in evaluation use.

Recommendation: EVA should consider adopting the recommendations on utility in Chapter 1.