

## Chapter 2

# Evaluation systems and governance

*This chapter explores the institutional arrangements and environments for evaluation across the DAC Network on Development Evaluation. The chapter highlights recent changes to evaluation systems since 2010, such as the creation of new evaluation entities and the integration of ministries of development co-operation with ministries of foreign affairs and trade. The chapter outlines the human and financial resources of evaluation units and examines evaluation policies as well as the independence of evaluation systems. The data and analysis is based on a survey of network members, document reviews, and interviews with evaluation units of DAC members and multilateral development banks.*

Evaluation systems are set up to provide regular information on the effectiveness of development co-operation interventions as a whole and on specific contributions of the various stakeholders involved in co-operation. Evaluation systems are established in an institutional environment that encourages knowledge sharing about the outcome of operations, as well as practical guidance and recommendations in order to improve operations. The institutional environments of development co-operation agencies and multilateral organisations are evolving and subject to political, technical and financial changes.

Evaluation systems need in some instances to respond to demands with inherently in-built tensions. Such tensions are apparent in the requirement for systems to promote both accountability and facilitate learning. Evaluations are key to providing the basis for holding partners to account for their interventions. Equally important, evaluations analyse successes and challenges, and set benchmarks for partners to learn from and to integrate into their future activities (OECD, 2010). Both demands coexist to varying degrees and force actors to develop practical systems that cater for them. Another example of stress lies in the demand that evaluations are independent from operations and management decisions. This demand is based on the independence that is required for impartial and credible assessments. At the same time, evaluation systems need to be pragmatic and effective, providing a link between the decision makers and operational units of the agencies.

Therefore, this review took stock of the variety of evaluation systems that EvalNet members have set up within their administrative and institutional structures. Some members have opted for centralised evaluation systems; others have promoted a strong decentralisation of the functions. Some agencies and organisations try to keep a balance. In recent years, some organisations have undertaken reforms in their evaluation system, which will be described. As a follow up of the previous review in 2010, this review has looked at the changes in evaluation policies, and furthermore, updated the financial and human resources dedicated to the evaluation function. The section goes on to discuss the issues of accountability and learning, as well as the independence of evaluation units, given their relevance for policy decisions and for the design of administrative structures.

### **Institutional setup of evaluation units**

The following section discusses some of the significant changes that have taken place during the last five years, and the location of evaluation systems within the administrative setup in order to understand the context in which evaluation work takes place. The questions posed in the survey sought to identify the drivers which determine the location of evaluation units within the administration and the position of evaluation systems on the spectrum of centralised and decentralised approaches. The section closes by looking at members' views of the respective advantages and disadvantages that the institutional setup may have, and about the potential factors that drive change.

## Recent organisational changes in evaluation systems

In recent years a number of agencies have undertaken reforms of the systems that support the delivery of development co-operation. In terms of the evaluation function, this has manifested in the establishment of bodies to improve the co-ordination of evaluation across a complex network of agencies, and to provide independent oversight of evaluation across the development co-operation sphere.

### **Creation of overall independent evaluation entities**

The German Evaluation Institute for Development Cooperation (DEval) and the UK Independent Commission for Aid Impact (ICAI) are examples of overall independent evaluation entities that have been established over the last six years to improve the influence of evaluations, the quality of strategic evaluations, and the effectiveness and impact of development assistance.

In 2012, Germany founded the German Evaluation Institute for Development Cooperation (DEval) in response to the recommendations by an external evaluation of the German evaluation systems. The evaluation found that there was a lack of coherence within and between the German systems, which were considered to be too fragmented. The most important recommendations were to define a clear division of responsibilities and labour between the different evaluation entities, and to increase the number of strategic evaluations, which are meant to enhance the steering capacity of the Ministry. The overriding goal of DEval is to independently evaluate the performance of German development co-operation interventions. DEval has reportedly already increased the influence of evaluation in the German development agencies, fostered greater visibility of the evaluation system, and improved the quality of strategic evaluations.

In the UK, the Independent Commission for Aid Impact (ICAI) was created in 2011 to scrutinise UK development assistance. ICAI carries out independent reviews of development cooperation programmes and issues related to the delivery of UK aid. Its mandate differs from the NAO (National Audit Office), which undertakes separate reviews/audits while ICAI reviews performance based on DFID evaluations, as well as carries out independent assessments of issues that affect aid delivery. ICAI reports directly to the International Development Committee in Parliament, and as such provides a line of accountability to the UK taxpayer. The creation of ICAI responded to the wish of the government to improve the effectiveness and impact of development assistance being provided across all departments.

The Swedish Government set up the Expert Group for Aid Studies (EBA) in 2013 to “continuously evaluate and analyse Sweden’s international development co-operation to improve development co-operation and build up a long-term, high-quality knowledge-base”. EBA benefits from what is labelled a “double independence”. It works independently from the government, chooses which issues it will evaluate, and to which entity the evaluations will be reported.

The French government had envisaged a similar move towards creating an overall independent entity, but the fact that the evaluation system depends on three different ministries complicated the decision.<sup>1</sup> Instead, the government opted for an Observatory, presided over by a Member of Parliament, providing advice on evaluations in the development co-operation domain only. The Observatory provides the communication link with stakeholders and civil society. It has a more consultative role than the independent ICAI and DEval.

The European Commission adopted the “Better Regulation Package” in 2015 to ensure that EU actions are more effective.<sup>2</sup> This package includes important policy and procedural changes to improve how it functions in many domains, including evaluation. It provides for the systematic publication of roadmaps that describe the problem to be tackled and the objectives to be achieved, feedback and consultation processes, and improved impact assessments, which collect evidence to assess if future EU action is justified and how such action can best be designed to achieve the desired policy objectives. A Regulatory Scrutiny Board (RSB) consisting of three EC Commission members and three external members examines all impact assessments and some major evaluations.

### ***The integration of evaluation for development co-operation***

A different kind of reform took place in New Zealand, Australia and Canada. The integration of the semi-autonomous unit of New Zealand Aid into the Ministry of Foreign Affairs and Trade (MFAT), the Australian Agency for International Aid (AusAID) into the Department of Foreign Affairs and Trade (DFAT), and the Canadian International Development Agency (CIDA) merger with the Department of Foreign Affairs, International Trade and Development (DFAIT) represents a new trend of combining the foreign policy, trade and development agenda. In these cases, the evaluation functions continue to be focused on development assistance. However, more recently, the Evaluation and Research team (MFAT New Zealand) has begun to engage with the wider Ministry to incorporate evaluation into its new strategic framework and four-year plan and to help establish an evaluative culture across the wider Ministry.

The MFAT New Zealand evaluation team is responsible for managing and delivering the strategic (sectoral, thematic, programme, policy and practice) evaluations. They are also responsible for the implementation of good evaluation practice, disseminating evaluation, research and results and building an evaluative culture. The Evaluation and Research team also provides technical advice and support to activity managers who are responsible for managing decentralised evaluations that are outsourced. The evaluation function of MFAT New Zealand is supported by an independent Evaluation and Research Board and overseen by the Deputy Secretary for International Development.

The Australian Government took the decision to integrate AusAID into DFAT with the goal of promoting greater coherence between Australia’s aid, foreign policy and trade efforts. The Office of Development Effectiveness (ODE) was originally established within AusAID and retained by DFAT as an operationally independent unit to measure and report on the effectiveness of the Australian development assistance programme. ODE’s position within DFAT provides it with direct access to aid management systems and personnel, which helps ensure mutual understanding of the work performed. The work of ODE is subject to the external oversight of the Independent Evaluation Committee (IEC), helping to ensure ODE’s credibility and independence.

In Canada, CIDA and DFAIT were integrated to create the Department of Foreign Affairs, Trade and Development (DFATD), now called Global Affairs Canada (GAC). The respective evaluation units have been maintained to fulfil their complementary functions. The Development Evaluation Division (DED) addresses programme effectiveness, programming tools, and priority and policy setting. DED also provides technical advice and quality assurance to programme branches and promotes organisational learning.

### ***The roles of centralised and decentralised evaluation units***

One of the drivers of the structural changes is the changing configuration of the evaluation function. Centralised evaluation units are based at the organisation’s headquarters and conduct evaluations (independent for the most part), and/or oversee or

validate decentralised evaluations. Most also provide technical assistance and training for operational units. In the case of decentralised evaluations, operations units that may be located at the organisation's headquarters or in partner countries, commission and oversee the evaluation. The evaluation can be carried out by the operations units themselves (self-evaluations) or be commissioned to external experts.

The majority of organisations have centralised evaluation units that are responsible for policy formulation; varying degrees of oversight, particularly of external evaluations; capacity building; and technical assistance for all forms of evaluations, specifically to decentralised units. Thirty per cent of survey respondents adopt a solely centralised approach to evaluation while 15% adopt a decentralised approach. Fifty-five per cent of organisations combine centralised and decentralised approaches. Among the multilateral organisation, five (62%) combine the centralised and decentralised evaluation functions and two (25%) undertake evaluations solely through centralised units.<sup>3</sup> Several evaluation units, including African Development Bank (AfDB), DEval, IEG and Policy and Operations Evaluation Department (Netherlands) (IOB) also conduct their own evaluations.

There are many advantages associated with the decentralisation of evaluations to locations close to interventions. According to the perception of the members, evaluations become more cost-effective and evaluation work performed closer to the intervention produces knowledge and information that can contribute to capacity building at local levels and within partner institutions. The survey and interviews suggest that information emanating from decentralised evaluation units caters for requirements at project and programme levels, while also being relevant to strategic decision making.

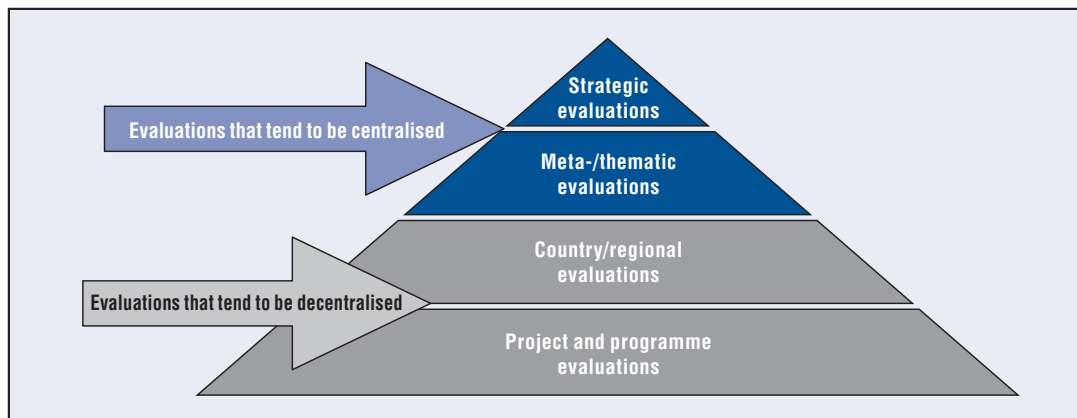
On the other hand, when evaluations are managed centrally, it may be easier to ensure that the quality is consistent as well as to ensure the coverage of evaluations. Furthermore the need for extensive support systems is decreased. Just under half of the survey respondents indicated that they have adequate capacity (in terms of human and financial resources) to provide training, support, advice and technical assistance to decentralised evaluations.

The decentralisation of evaluation in DFID has resulted in some significant adjustments. The introduction of a cadre of professional evaluation advisers that supports decentralised evaluation facilitates the ownership of findings at an operational level. However, more recently, due to resource constraints, there has been a decrease in evaluations advisers: they are either part-time (split between country offices) or perform more than one role, combining evaluation with results management and statistics (based on the interviews conducted for the survey).

To support the evaluation function and work programme, MFAT (New Zealand) has established a range of mechanisms to provide independent technical, subject-specific and cultural experts to support its work. An independent Evaluation and Research Board also gives advice and support to the Aid Programme on evaluation policy, practice and function and endorses the multi-year evaluation programme.

It is challenging to ensure the coverage of evaluations that address the various organisational needs, i.e. strategic, thematic down to programme and project, as confirmed by the Review of Embedding undertaken by DFID (DFID, 2014). This is due to the autonomy provided to the operations units to determine if an evaluation is a necessary. This provides increased flexibility for operational managers (country offices and programme teams) to determine whether or not to perform an evaluation on specific programmes and projects, in line with their Country Strategies, EvD guidance and administrative and planning requirements for developing the business case for new initiatives. This has affected the ability of the central unit to use the decentralised evaluations as the basis for the synthesis work that they are responsible for producing (Figure 2.1).

Figure 2.1. Typical types of evaluations that are centralised and decentralised



However, for organisations whose requirement to undertake decentralised evaluations is based on project value and completion, such as MFAT New Zealand, the coverage of evaluations was not highlighted as a significant issue.

### Co-ordination

An obvious challenge to administering evaluations lies in the necessary links between centralised and decentralised units. The vast majority of members indicate that links are weak or moderate; only five out of 25 respondents describe the links as stronger and formal. Most centralised units conduct quality control for the terms of reference of decentralised evaluations; in some cases, they elaborate the policy for decentralised units, and provide guidance, tools and technical assistance. In some other cases the communications are more frequent with informal discussions and written feedback.

The review shows that the location of evaluation units is important for the substance of evaluations. Most of the agencies stress the prime need for evaluation units at the central level, close to management decision and policy making. But agencies are aware of advantages of evaluation by decentralised evaluation units, alongside operations units. Recently, agencies have moved to strengthen the links between centralised evaluation and decentralised evaluation agents.

In practice, agencies align in different ways at the various structural levels. In Global Affairs Canada for example, the evaluation unit has a team dedicated to provide support to those carrying out decentralised evaluations. Formal links exist also, such as contracts signed between the evaluation team and the centralised unit, with close and direct co-operation during the whole evaluation process via meetings and e-mail communication. The Finnish Ministry of Foreign Affairs strengthens the link by getting a closer insight into the needs of the decentralised evaluators. Every two years, the central evaluation unit commissions meta-evaluations, which are based on decentralised evaluations.

However, for institutions that predominantly rely on decentralised evaluations, co-ordination systems are all the more important. Of the six institutions that responded that this is the case for them, including *Deutsche Gesellschaft für Internationale Zusammenarbeit* (GIZ) (Germany), MFA Poland, MFA Slovakia, SIDA, USAID, and the US Department of State, the role of the central evaluation unit is reportedly to be primarily to provide technical

assistance and quality assurance to the decentralised units in undertaking evaluations. SIDA, for instance, confirmed that the links between the central evaluation unit and decentralised units are in fact weak but that this is being improved.

There are also relevant examples of inter-agency co-ordination. Among the French agencies, there is already a markedly close co-ordination among the three evaluation units. The heads of units meet frequently, at least every three months. They organise common working groups and trainings, and set up active cross participation in steering committees. The evaluation units discuss work plans to define joint evaluations, and submit a common report to Parliament.

In summary, the review proposes two conclusions in relation to the institutional setup of evaluation units. Although it is too soon to measure the success of specialised oversight/scrutiny bodies, their creation testifies to the need to provide for robust evidence on accountability by independent scrutiny of development assistance and co-ordination of complex evaluation functions related to ODA. This may point to the view that where accountability is taken care of by the oversight/scrutiny bodies, evaluation units are able to focus on learning. This does not imply, at least not yet, that the accountability dimension will be addressed by specialised autonomous entities. The twin objectives of evaluation, to provide evidence for accountability and the basis for the institution to learn, remain entwined.

The majority of organisations make use of a hybrid of centralised and decentralised evaluation, increasing cost-effectiveness, relevance and ownership of evaluations for operational units, and supporting the development of capacity in partner institutions. However, ensuring the quality, coverage (of sectors, themes or strategic issues) and consistency of decentralised evaluations is challenging, affecting the usability of evaluations for synthesis purposes. In some cases, central evaluation units have set up extensive support systems, providing assistance at all levels of the evaluation process, from the review of the terms of reference, to the templates for quality control and peer reviews. Meta-evaluations help in pointing out weaknesses and the places where support is needed, and as such improve the systems.

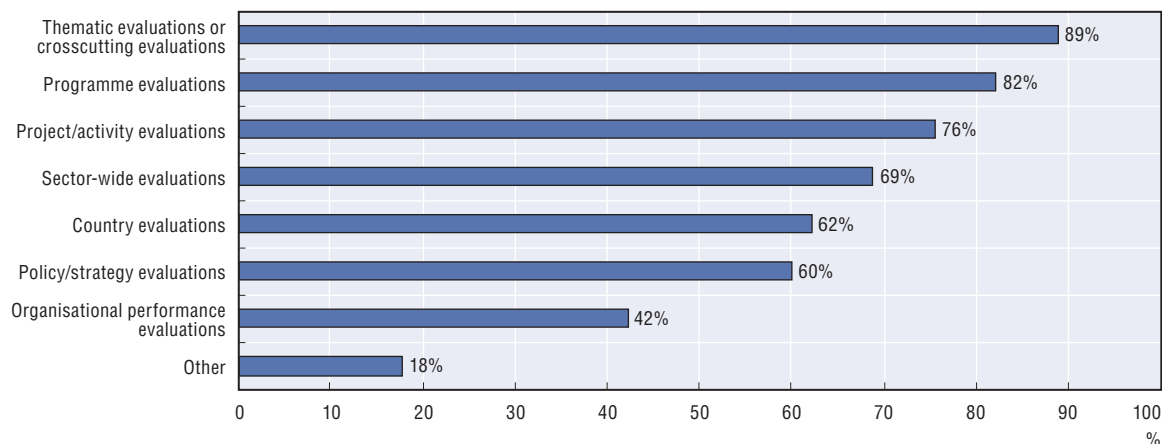
## Financial and human resources related to the evaluation system

This section outlines the types and volume of evaluations being implemented, appraises the changes that have been made in human and financial resources committed to evaluation systems since 2010. The section therefore starts with an overview of the different evaluation activities that are undertaken across the member organisations.

### **Types of evaluation**

Figure 2.2 describes the types of evaluations carried out by both multilateral organisations and bilateral agencies. Similar to the patterns identified in 2010, almost 90% of members undertake thematic evaluations. Although there is an increased emphasis in member policies and planning on strategic priorities, the use of policy/strategy evaluations appears to have become less common (from 92% in 2010 to 60% in 2016). Where 49% of institutions were performing project/activity evaluations in 2010, 76% are now undertaking them.

Figure 2.2. **Types of evaluation**  
Percentage of responding evaluation units that conduct

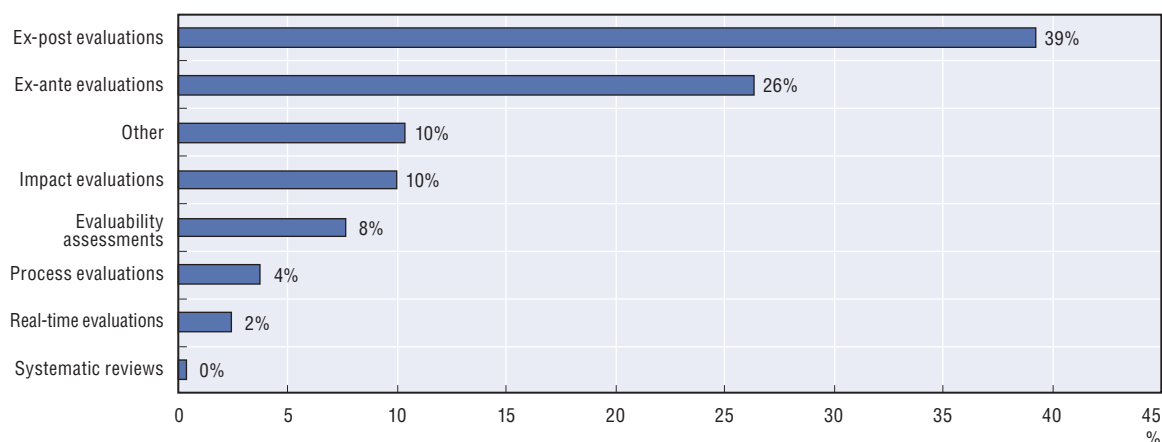


Source: Online survey completed by EvalNet members for the purposes of this report, <http://www.oecd.org/dac/evaluation/reviewofevaluationsystems.htm>.

The total number of evaluations undertaken during the last five years is reported to be 5 916 (reported by 6 multilateral organisations and 27 bilateral agencies).<sup>4</sup> Although the 2010 review found that the trend was to move away from ex-post evaluations, 39% of evaluations performed in the past five years are reportedly ex-post and 26% ex-ante. Over 580 impact assessments have been performed during the period (500 by GIZ) and 452 evaluability assessments (120 by the Inter-American Bank and 250 by Korea International Cooperation Agency [KOICA]).<sup>5</sup>

Out of the 1 557 ex-ante evaluations performed during the past five years, 1 262 were performed by Japan International Cooperation Agency (JICA) and 200 were implemented by KOICA, reflecting the emphasis in their policies on both types of evaluation. Based on these aggregate figures, an average of 26 reports are produced by each evaluation department per year, similar to the average of 24 in 2010, however due to the variances in terms of the intensity and frequency of evaluations, this may not be particularly illustrative.<sup>6</sup>

Figure 2.3. **Volume of evaluation**  
Percentage of evaluations undertaken by type during the past five years



Source: Online survey completed by EvalNet members for the purposes of this report, <http://www.oecd.org/dac/evaluation/reviewofevaluationsystems.htm>.



### Human and financial resources

The comparison of spending and resources allocated to evaluation across member organisations is challenging because of the very different institutional structures, the different evaluation activities that are undertaken and the different methodologies and reporting systems used. Therefore the following section focuses on the trends that are apparent in relation to the resource allocations to the evaluation function and the perceived impact, as described by survey respondents.

Obtaining data on financial resources and average number of evaluations performed was challenging because a range was usually provided. The resources attributable to centralised evaluations were reported, but due to the fact that the budget for decentralised evaluations is taken from the programme or project budget, this data was not available. Evaluation units also undertake other activities such as research and knowledge management, which is indistinguishable from the budget allocated specifically for evaluation.

These challenges aside, the findings of the survey provided some insight into recent changes in the human and financial resources allocated to the evaluation function of member organisations. EvalNet members employed 664.5 staff members in 2016.<sup>7</sup> The average number of full-time evaluation staff has not changed significantly since 2006 (13.8 in 2006 to 14.1 in 2016) (OECD, 2006). Forty-two per cent of organisations reported that their staff had increased, whilst 11% (four bilateral agencies and one multilateral organisation) had decreased their staff.<sup>8</sup>

Table 2.1. Summary of changes over the past five years in budget and staff in multilateral organisations and bilateral agencies

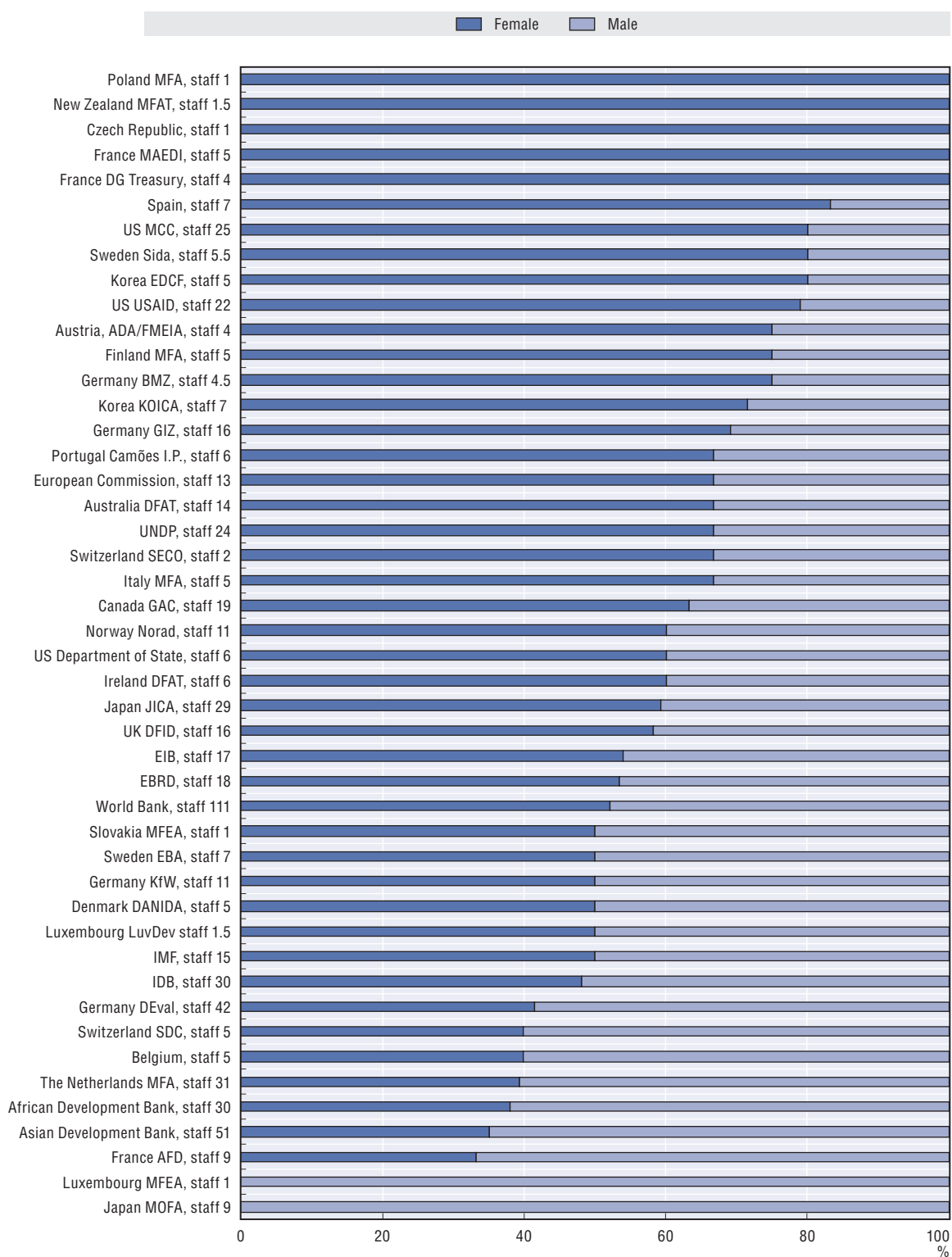
Changes	Changes to budget		Changes to staff	
	Multilateral organisations (number of)	Bilateral agencies (number of)	Multilateral organisations (number of)	Bilateral agencies (number of)
Increased	3	12	5	14
Decreased	2	8	1	4
Stayed the same	2	14	2	18

Note: Some respondents did not provide complete information on budget or staff. The data in this table does not include non-responses.

Source: Online survey completed by EvalNet members for the purposes of this report, where information was available<sup>9</sup>, <http://www.oecd.org/dac/evaluation/reviewofevaluationsystems.htm>.

From the survey, the number of staff working in the evaluation units of bilateral development organisations varies from one person working halftime to 42. Eight out of 37 bilateral agencies have two or less staff in their evaluation units. The staffing ranges from 13 to 111 for multilateral organisations, reflecting the broader scope and budget associated with their activities. Disaggregated by gender, the average percentage of female staff is approximately 62% and 38% male. This is a similar proportion compared to the 2010 data.<sup>9</sup> If administration staff are excluded, the proportion is more balanced: 56% female and 44% male. The proportion of women and men in the role of Head or Director of Evaluation is almost equal (46% of the staff holding these positions are women).

Figure 2.4. Gender disaggregation of employees working in centralised evaluation units  
(professional and management staff)



Note: The number of staff reported in the vertical axis is the number of full-time equivalents.

Source: Online survey completed by EvalNet members for the purposes of this report, <http://www.oecd.org/dac/evaluation/reviewofevaluationsystems.htm>.

During the last year, the budget of evaluation units of 34 bilateral agencies ranged between EUR 30 000 and EUR 20 million. The median budget is EUR 1.05 million and average EUR 2.76 million. Only five bilateral agencies reported a budget lower or equal to EUR 100 000. Eight multilateral organisations reported a budget for their evaluation units of between EUR 3.65 million and EUR 30.6 million.

Table 2.2. Summary of evaluation resources<sup>10</sup>

Network overall	Average 2010	Average 2015	2015		
			Median	Maximum	Minimum
Spending on evaluation (EUR)	4.08 million <sup>11</sup>	4.1 million <sup>12</sup>	1.5 million	30.6 million	30 thousand
Number of staff in central evaluation unit (total staff)	19	14	7	111	0.5
Number of evaluations per year incl. ex-ante evaluations	24	35	8	416	1.0
Number of evaluations per year excl. ex-ante evaluations	n.a.	26	8	163	1.0
Evaluations completed per year per staff member (total staff) <sup>13</sup>	1-2	1-2	1.12	15.9	0.13
<b>Bilateral agencies</b>					
Number of staff in central evaluation unit (total staff)	n.a.	9	6	42	0.5
Budget for evaluation	EUR 1.92 million	EUR 2.76 million	EUR 1.05 million	EUR 20 million	EUR 30 thousand
Central unit budget compared to the development budget	0.16%	0.20%	0.10%	1.00%	0.03%
<b>Multilateral organisations</b>					
Number of staff in central evaluation unit (total staff)	n.a.	34	24	111	13
Budget for evaluation	EUR 7.76 million	EUR 9.96 million	EUR 7.90 million	EUR 30.6 million	EUR 3.65 million
Central unit budget compared to the administrative budget	1.38%	1.19%	1.15%	2.00%	0.16%

Source: Online survey completed by EvalNet members for the purposes of this report and verified member profiles, <http://www.oecd.org/dac/evaluation/reviewofevaluationsystems.htm>.

The most significant change that has occurred since 2010 is the decrease in the average number of staff. Despite the increase in the average budget for multilateral organisations, there is a decrease in the central evaluation unit budget compared to the administrative budgets.

The central evaluation units of 24 organisations are funded through a budget that is separate from the operational budget.<sup>14</sup> This does not necessarily mean that the central evaluation unit benefits from a budget that is fully independent from senior management. The multilateral development banks, for example the African Development Bank (AfDB), the Inter-American Development Bank (IDB), and the World Bank, are allocated budgets on a distinct budget line over which the senior management has no influence. Most bilateral agencies have budget lines that are separate from the operations units, but are still dependent on senior management or ministerial decision to approve and allocate budgets. MCC presents another approach in that resources for evaluations are included in compacts (projects) at their inception for data collection during project execution; resources from the agency's due diligence budget are provided for the implementation of evaluations, specifically the recruitment of independent evaluators.

The Special Evaluation Office (SEO) (Belgium) undertakes only centralised evaluations, and receives two basic allocations (separated from operational budget): the first one covers SEO's operational expenses and ordinarily amounts to at least 0.10% of the ODA to be evaluated, and the second allocation contributes to the funding of joint evaluations. During the survey some agencies pointed out the need for separate budget lines, and for specific and predictable budgets for centralised evaluations.

Decentralised evaluation is funded from operational budgets for thirteen member organisations, approved by senior management. In some cases, the programmes and projects have already included evaluation funding in their costs, or decentralised evaluations are covered through extra-budgetary funding.<sup>15</sup>

How the agencies responded to a change in resources is even more important than the change itself. The survey responses showed that agencies benefitting from an increase in financial and/or human resources generally used these resources to further develop capacity or quality assurance, to expand their scope of work, or to improve knowledge management.

Where financial and/or human resources remained the same, two agencies indicated that they carry out fewer evaluations while keeping high quality standards; the World Bank notes that the unit has increased internal efficiencies using measures such as a department restructure and the harmonisation of portfolio analyses.

Among the agencies that had to face a decrease in financial and/or human resources, the Ministry of Foreign Affairs of the Netherlands became more selective in conducting evaluations; more critical about the necessity for impact evaluations; and delegating responsibility for decentralised evaluations in combination with an enhanced helpdesk function. The evaluation unit raised concerns about the fact that reduced staff and budgets may affect the quality of evaluations. SIDA decreased the number of centralised evaluations, although it expects an increase in staff of up to five evaluators in 2016. It intends to spend more on resources to improve the quality of decentralised evaluations and draw on them for synthesis and lesson learning. Danish Development Assistance, Danida, on the other hand, conducts slightly fewer and less costly evaluations than before.

Decreases in budgets for operational units have also affected decentralised evaluation. DFID country offices have to prioritise evaluations to be undertaken and consider whether an Evaluation Adviser is a necessary budget expense. Some evaluation advisers therefore cover more than one country office. The production of synthesis reports by the DFID Evaluation Unit is affected by the reduced number and variety of evaluation reports produced, due to budget restrictions.

On the whole, agencies responded to changes in resources affecting the evaluation system, both centralised units and decentralised operation, by focussing on learning and making an effort to increase the relevance and strategic impact of the evaluations commissioned. Thus, agencies that benefit from increased resources do not increase the number of evaluations, but rather focus on issues likely to bring about new directions, for example meta-evaluation, innovative methodology, and impact at strategic levels.

## Evaluation policies

The prevalence of evaluation policy documents is an important indicator of the recognition of the priority placed on evaluation as a core function of most organisations. This section reports on changes in evaluation policy documents, highlighting perceived needs and providing examples of a shift in policy for a number of organisations.

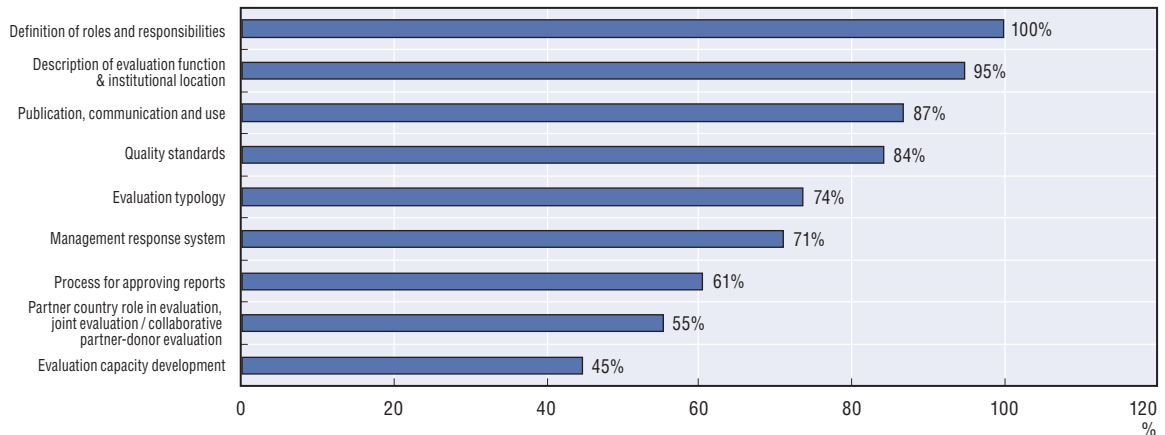
Between 2010 and 2015, there was a notable increase in the number of agencies with a policy document defining their evaluation activities. In 2015 38 out of 46 member agencies (83%) had a policy document, compared to 28 out of 40 agencies (70%) in 2010. Furthermore, those that do not have a policy have sets of guidelines and methods. Five member agencies are in the process of preparing and approving evaluation policies.

### Policy content

The majority of policies describe the evaluation function and institutional location, roles and responsibilities, quality standards and publication, communication and use of evaluation reports and findings

Figure 2.5 below provides a summary of the content of members' evaluation policies.

Figure 2.5. Content of evaluation policies



Source: Online survey completed by EvalNet members for the purposes of this report, <http://www.oecd.org/dac/evaluation/reviewofevaluationsystems.htm>.

Most of the members surveyed state that their evaluation policies are expected to be revised or updated within the next five years. The aims cited include: to increase evaluations at a strategic level; to be more effective in influencing organisational decisions; and to increase the scope of evaluations.

Despite the achievements that have been made in strengthening the alignment between policy and strategic priorities, there is still room for improvement. On the survey question “whether the policy is congruent with organisational strategy”, 14 respondents consider their evaluation policy to be very congruent with organisational strategy (out of which only one was a multilateral organisation); 22 found it moderately congruent (all bilateral organisations); and four somewhat congruent (out of which only three were multilateral organisations). Some respondents would like the evaluation policy to create more ownership within senior management.

Likewise, the relevance of the evaluation policies in relation to actual needs may be furthered. Regarding the responses to the survey on “how to make the policy more relevant to organisation’s needs”, the main arguments presented include: underlining partner participation; better articulating relationships between evaluation and other related functions such as monitoring, review and research; and balancing the accountability and learning roles of evaluation.

The evaluation policy may need to undergo a change to reflect changes in development policy. For instance, one survey respondent replied that the policy had to be more relevant to organisations’ needs: “To take account of the (now) Department-wide remit of the evaluation function, and the Department’s new Global Foreign Policy, new development assistance policy and the Department’s associated Statement of Strategy (including the roles of business units and their partners in relation to evaluation).”

Among the agencies some distinguish clearly between an evaluation policy and an evaluation strategy. A respondent reply to the survey question on how to make the policy more relevant to the organisation's needs: "I don't think this is a question of having an appropriate policy but of having an evaluation strategy. Politics might change every four years after elections, changing an evaluation policy so frequently seems neither necessary nor appropriate". The African Development Bank's revised policy no longer includes a list of Independent Evaluation products and procedures, which are instead discussed in the Independent Evaluation Strategy.<sup>16</sup> Some other respondents view the policy to be more of a set of guidelines: "We need to include in the policy document how to put our policy into practice and the quality standards. There is also a need to give more instruction on how to use the results of our evaluations."

Comparing with the previous review findings, there are only a few noticeable changes. Whereas there seems to be slightly less emphasis on the need to include evaluation capacity development in the evaluation policy document (45% of the actual policy documents include it, against 50% in 2010), more evaluation policies are including directions on the role of partner countries in evaluation and provide a mandate for joint evaluation and collaborative partner-donor evaluation (55% against 50% in 2010). This is actually a growing tendency, with more institutions envisaging including this.

Significant evaluation policy changes have been made within US agencies, the World Bank Group and EBRD. For example, the updated policy of the US State Department (2015) includes the requirement to publicly post all evaluations within 90 days of their completion. Full evaluations are posted except in the case of sensitive or classified issues where only a two to three page summary is posted. The positive outcome of this new policy is that the requirement to publicly post all evaluations reportedly had a definite impact on the quality of evaluations. This resulted in a better and more careful selection of evaluations to perform and a better quality of evaluations in terms of methodology, validation of data and evidence, and recommendations.

A review of the number of evaluations submitted to USAID's Development Experience Clearinghouse (DEC) has shown a decrease in evaluations from nearly 500 conducted in 1994 to approximately 170 conducted in 2009, despite an almost three-fold increase in programmes.<sup>17</sup> Over that period, the agency's evaluation activities have been subject to both internal and external critiques regarding methodological quality, objectivity, access to evaluation findings, and use of evaluation recommendations for decision making. In January 2011, a new policy entailed important changes:

- The budget allocated to evaluation activities is equivalent to 3% of all programme funds.
- There is now an M&E specialist contact person responsible for each country bureau.
- All evaluations are posted on their website in full.
- Strict requirements for impact evaluations: pilot programme as well as for innovative development interventions.
- Evaluating the performance of all large programmes i.e., equal or greater than the average size of projects in countries.

Since then, an average of more than 200 evaluations is carried out each year, with more than 1 000 evaluations completed since 2011.

The current evaluation policy of MCC was established in 2012. This policy will be updated with more precise guidelines and procedures. These are already applied but have not yet been formalised in the policy itself.

During the 2013 revision of the EBRD evaluation policy, the document was abbreviated significantly from over 70 to 12 pages in order to separate guidance on process from policy. The policy therefore focuses on the basis for the evaluation function at EBRD, including: the coverage and objectives; purpose and principles of evaluation; roles and responsibilities; procedures; access to information and distribution of reports; and special provisions (i.e. role of Chief Evaluator and management of resources). The aim of the 2013 strategy was to integrate evaluation as a core and valuable function within the Bank. At the time, evaluation was considered as an add-on function. The emphasis of the policy also shifted towards a greater balance between accountability and learning.

To summarise, the number and percentage of organisations that have a policy document to guide evaluation work have increased, but the purpose of the document differs from providing a set of guidelines or policy, to being a strategy. Member organisations are placing an increased focus on the relevance of evaluations in relation to the organisation's needs, such that the policy better articulates relationships between evaluation and other related functions such as monitoring, review and research; and balancing the accountability and learning roles of evaluation. An increasing number of agencies view partner involvement as a guiding principle and have either already included this in the policy document, or plan to do so. The evaluation policies are being revised regularly, e.g. in light of changes in structures, or based on new focus areas, e.g. the Sustainable Development Goals (SDGs).

## Accountability and learning

The twin objectives of accountability and learning are key to any evaluation system. Since the beginning of independent evaluation practice, the concepts have had contrasting perspectives with shifting weighting towards one or the other perspective. The shifting perspective is summarised by the following quote: “One theory suggests that accountability impedes or undermines learning; and as long as evaluation tries to do both, it will not be effective. The other perspective suggests learning and accountability are mutually reinforcing. In particular, learning is derived from an understanding of success and failure to determine how to be more effective in future. In this case, accountability is the foundation for learning” (Heider, 2016).

Accountability is described by the OECD DAC as follows (OECD, 1991):

The accountability notion of evaluation... relates to the developmental results and impact of development assistance. It is distinct from accountability for the use of public funds in an accounting and legal sense, responsibility for the latter usually being assigned to an audit institution. Information about the results of development assistance should be provided to the public and their leaders in both donor and recipient countries. This may require a careful, even legally defensible, demonstration that the work is consistent with the contract terms. [Further,] accountability in development may refer to the obligations of partners to act according to clearly defined responsibilities, roles and performance expectations, often with respect to the prudent use of resources. For evaluators, it connotes the responsibility to provide accurate, fair and credible monitoring reports and performance assessments. For public sector managers and policy makers, accountability is to taxpayers/citizens.

The changing weight on either the accountability or the learning aspect can be illustrated by the history of the World Bank, which was the first development institution to establish an independent evaluation unit. The evaluation unit was set up in 1974 and reported to the Board of Directors. At that time, accountability was the main focus of evaluation. In the 1990s, this focus received severe criticism because operations suffered setbacks and failures due to the emphasis on accountability, at the cost of performance management.<sup>18</sup>

In 2015, an external review of the World Bank Group (WBG) Independent Evaluation Group (IEG) was conducted at the request of the Committee on Development Effectiveness (CODE) (Weiss et al., 2015). This review recommended an increased emphasis on learning, and urged IEG and WBG management “to design, build, and continue to strengthen an overarching evaluation, learning, and accountability policy and system”. (Weiss et al., 2015) This system would better enable IEG and CODE to play their respective roles in the cycles of learning, course corrections, accountability, and continuous improvement for the World Bank Group to achieve its development goals.

Beyond the World Bank, the debate also helped to progressively form the concept of institutional learning as a balance to accountability in development organisations. Today, most evaluation units in their policy documents express the view that evaluation pursues both objectives: learning and accountability. The survey largely confirmed this view. Responding to the question of “what is the major purpose of evaluations” the majority of member organisations (26 out of 46) said that they favoured a balance between the two concepts. Six organisations indicated that they viewed the learning part as having slightly more importance, 12 organisations viewed learning as being the sole objective of evaluation. Of multilateral organisations, four proposed a balance between accountability and learning, one favoured the sole objective of learning, and one agency had a slight propensity towards learning over accountability.

The following responses represent a typical sample:

- “Learning is important because the knowledge produced by evaluations is precious and valuable and this needs to be shared for the organisational benefit, but having learning as the sole purpose of evaluation wouldn’t require an evaluation function, and especially not an independent evaluation function”.
- “We are also aware of (and try to manage) the fact that evaluations can and should be structured depending on the purpose (accountability or learning) and that combining both in the same evaluation may give sub-optimal results”.
- “In my view accountability (in particular for an evaluation function that is ‘in-house’) fosters ‘hard top-down management’ which produces procedures and a work culture which produces a lot of “rituals” (i.e. documents and processes that are meant to please top management and those who demand accountability, but with limited use for the operational staff). A learning organisation has the potential of doing the opposite, i.e. to foster a work culture where procedures and methods are meaningful for the operational staff”.
- “Both are very important, and I would say that a balance is critical. There are few other accountability mechanisms for IFIs, so this is a critical role. Learning is also important, particularly given the unique access to information that evaluation function has”.

In summary, the survey results and interviews reflect perspectives on the balance of the two objectives, or rather a preference for one of the objective over the other. The issue of independence of evaluation units, which will be discussed in the next section, is often linked to their accountability dimension.

## Independence of evaluation system

This section deals with the principle of independence as a function of evaluation systems that instils trust in the evaluation function as related to governance/oversight (Mayne, 2011).<sup>19</sup> In other words, duly independent evaluations are part of sound management practice. In this respect, independence reflects on three dimensions: organisational independence, behavioural independence, and independence from influences and conflict of interests.



The review asked about the significance of independence. Members' opinions about advantages and disadvantages of independent evaluation units were weighted in favour of advantages. The respondents believe that independence of evaluation results has an important impact on the uptake of results in subsequent decisions.

The advantages of independence that were highlighted by survey respondents primarily related to the increased credibility of evaluations, as perceived mainly by external stakeholders. The objectivity of the evaluation function conveys a commitment to accountability to the public. The political pressures and administrative burdens that are placed on operational units to apply a specific focus are not applied to the evaluation function. This detachment allows the evaluation function to present governing bodies or executive boards with open reflections and assessments that are not framed in furtherance of management's interests and agenda.

On the question "Do you feel that the degree of independence of the evaluation function is sufficient to ensure external credibility of evaluation?" the large majority of respondents from multilateral organisations (five against three) fully agree, while only 10 respondents from bilateral organisations fully agree, 20 agree moderately, and four agree to some extent. There is, no clear trend among the respondents from multilateral organisations on the extent to which independence is perceived as important in the update of evaluation results, while there is a clear weight of respondents from bilateral agencies that find it "important" or "very important".

There is clearly a balance that needs to be struck between independence and isolation. The reported disadvantages include: a lack of connection with operational units; a reduced sense of ownership of the evaluation results, specifically by management, affecting uptake; risk of being perceived as disloyal and untrustworthy or bringing unwelcome criticism that is potentially damaging to upholding confidence and corporate well-being; and the lack of contextual knowledge of those evaluating, necessitating that the evaluands spend significant amounts of time providing this background.

The organisational independence of the evaluation units is clearly described in the policies of the majority of member organisations, showing that it is clearly a priority. Respondents confirm that the principle is well anchored within the multilateral organisations. All but four bilateral agencies respond that the independence of the evaluation units is mentioned within their evaluation policy. In the cases where it not mentioned it is implied.

Reporting lines are an indication of independence: structural independence means that evaluation units do not report to their management, but to boards of directors. This prevents evaluation units from being controlled or influenced by decision makers, and enables them to have access to organisational information and to submit evaluation findings without interference from their management or from project or programme managers. If the evaluation unit is under the authority of an operational or administrative department or senior management than work plans and the preparation and implementation of evaluations may be subjected to internal reviews and controls by those who are being evaluated. This thereby puts at risk the objectivity of the function.

In fact, a considerable number of evaluation units report to operational management, which may give rise to conflicts of interests. The survey indicates that out of 38 bilateral agencies, 21 report to senior management or related departments; five to the Minister or Deputy Minister; two to the Parliament; four to their board of directors; three to the Secretary General of International Development (or Under Secretary); three did not respond. All evaluation units of multilateral development banks report to their Board of Directors

either directly or through their Committee for Development Effectiveness (CODE), except for EIB (reporting is to the Inspector General).

Nevertheless, evaluation units are confident that they are largely independent from operational sections. The overwhelming majority of respondents from bilateral agencies feel that they are fully, or for the most part, independent from the influence of operational staff. Only two staff feel “to some extent independent” and one “not at all independent”. Six respondents out of eight from multilateral organisations feel that they are fully independent from operational staff.

The assurance of being independent in cases where there is no formal organisational independence is related to safeguards that protect evaluation units from interference in those bilateral agencies reporting to their management. The review offers some examples:

- In USAID, although the evaluation unit reports to the Head of the Bureau for Policy, Planning and Learning, the Inspector General is the guarantor of independence for central evaluation units, which can if necessary, resort to it in the case of infringement of the independence of evaluation activities.
- In MCC, there is a strong perception of independence even if the evaluation unit is not structurally independent. All evaluations are performed by independent entities. The agency’s M&E policy provides for complete independence in terms of method, content, and dissemination of evaluation findings.
- In the evaluation department (IOB) of the Netherlands, independence is guaranteed through its autonomous role in programming and approval of the terms of reference; independence in the evaluation approach (methods and techniques), independence in selection of consultants (if part of the evaluation is outsourced), and reporting. The responsible minister sends the policy reviews and evaluations to parliament, with a policy reaction. All evaluations, policy reviews and studies are published, unless they were committed from the start as internal (organisation) advice.
- In the Ministry of the Foreign Affairs of the Czech Republic - Development Cooperation Department (MFA-DCD), though there is no structural independence, the selection, implementation, and reporting functions are carried out independently and prepared under the guidance of the Working Group of the Council for Development Cooperation, which, along with reference groups composed of independent members, provides some degree of impartiality.

In addition to reporting lines, the evaluation function may be impaired by personal career decisions of the head of evaluation. In this regard, the survey asked whether the recruitment process could create a bias against independence. On the question on whether the posts of heads of evaluation units are filled internally or externally, most respondents marked the question as “not applicable”. The remaining respondents were in favour of external recruitment of heads of units. On this basis, the modality used to recruit the head of evaluation is not considered to compromise independence.

In multilateral organisations, only two respondents (against six) replied that the Head of Evaluation Unit has the possibility of moving to another position within the organisation when they leave the role, while in bilateral agencies most of the members (30 against five) could do so. This points to the possibility that organisations need to safeguard independence in cases where the head of evaluation considers career changes. One may draw the cautious conclusion from the survey that neither the recruitment of candidates for positions of head of evaluation unit, nor internal career prospects of heads of units appear to impact the independence and impartiality of evaluation units.

Behavioural independence is defined as “the extent to which the evaluation unit is able and willing to set its work programme, produce high quality and uncompromising reports and to disclose its findings to the Board without management-imposed restrictions. Conflict of interest safeguards guarantee that current, immediate future or prior professional and personal relationships and considerations are not allowed to influence evaluators’ judgments or create the appearance of a lack of objectivity” (Picciotto, 2013). Although behavioural independence was not covered by the survey, some of the responses from the survey and interviews highlighted the importance of the principle. The suggestion was made that professionalism and constructive engagement is extremely important for enhancing credibility where structural independence is not strong.

Participation in senior management meetings allows heads of evaluation units to be informed of the organisational strategy and policies. While this can be done through the communication of minutes, physical participation also provides the opportunity to make evaluation better known, thereby contributing to development of the evaluation culture and becoming familiar with the issues at stake at senior levels. The trend is indeed that evaluation unit representatives participate in management meetings. The survey shows that over 80% of bilateral agencies include representatives of central evaluation units in management meetings. The routine participation in management meetings of evaluation staff is varied between the multilateral organisations.

Overall, survey respondents perceive the position of evaluation units to be relatively free from interference. On the question whether the staff “feels that they are able to take decisions regarding evaluation without interference from their supervisors?” 29 respondents from bilateral agencies indicated “always” or “for the most part”, whereas six of the respondents replied “to some extent”, and one “not at all”. Among multilateral organisations, the eight respondents replied “always” or “for the most part”. The degree of independence is perceived to be high in relation to evaluation planning and preparation, report writing and dissemination, and lower on the selection of evaluations.

In summary, the principle of Independence as a function of evaluation system, as well as part of the overall institution, is well rooted in the culture of EvalNet members. The advantages of independent evaluation clearly outweigh the disadvantages. In accordance with evaluation principles, the organisations in their entirety have to continuously work on reducing misalignments to integrate the evaluation function into operations. Evaluation unit staff are aware of aspects of independence: organisational independence, behavioural and freedom from interference.

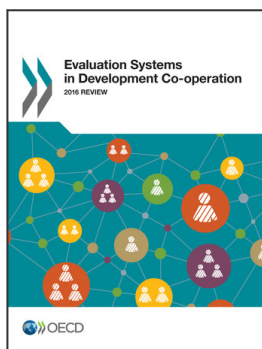
## Notes

1. AFD is under the umbrella of 3 ministries: Ministry of Economy and Finance; Ministry of Foreign Affairs and International Development; and Ministry of Immigration, Integration, National Identity and Co-Development.
2. European Commission, Better regulation: why and how, [http://ec.europa.eu/smart-regulation/index\\_en.htm](http://ec.europa.eu/smart-regulation/index_en.htm). (All websites were accessed during the period January – May 2016). Better regulation is about designing EU policies and laws so that they achieve their objectives at minimum cost. It ensures that policy is prepared, implemented and reviewed in an open, transparent manner, informed by the best available evidence and backed up by involving stakeholders. Better Regulation is partly based on the REFIT (Regulatory Fitness) Programme 2011-12, which was a rolling programme to keep the entire stock of EU legislation under review and ensure that it is “fit for purpose”, that regulatory burdens are minimised and that all simplification options are identified and applied.
3. The IEO of the IMF did not reply to this question in the survey.
4. Data obtained from online survey completed by EvalNet members for the purposes of this report. Excluding ex-ante evaluations this total is 4 359.
5. The individual respondent definition of the types of evaluation may have differed for the types of evaluations, e.g. impact evaluations (which may have included impact assessments) and ex-ante evaluations (which may have included project appraisals).
6. This includes 1 557 ex-ante evaluations and 452 evaluability assessments over the last 5 years, without which the average number of evaluations performed per evaluation department per year is 17.
7. Based on full-time equivalents employed by member organisations as opposed to number of individuals.
8. In the case of BMZ, the decrease was offset by a significant increase in staff at DEval due to the restructure.
9. Figure 2.4 provides an overview of the number of staff disaggregated by gender for all respondents. The number of staff that constitutes the basis for Figure 2.4 is greater than the number of full-time equivalents reported next to the name of agency/organisation as individuals were counted.
10. Data in this table are compared to the 2010 OECD review report. Assumptions have been made that the number of evaluations in 2010 were reported as they were in 2015.
11. Based on average USD:EUR exchange rate of 0.8 for 2010.
12. Based on reported budget in 2015.
13. The numbers are excluding ex-ante evaluations. The annual average including ex-ante evaluations is three to four per staff member.
14. Three organisations did not answer the question or were not explicit.
15. The extent to which changes in resources over the last five years have affected the evaluation work and activities was assessed in the survey with open-ended questions. No information is provided on decentralised evaluations given that centralised units rarely have this information.
16. The ADB’s Evaluation Policy currently in force was approved in 2007. In 2012, IDEV submitted a proposal for a revised policy which made significant changes to the 2007 policy. The proposal has been discussed on several occasions by the Board of Directors and has undergone adjustments, but the revised policy has not been approved by the Board yet.
17. USAID’s Development Experience Clearinghouse is the largest online resource for USAID-funded technical and programme documentation, with more than 155 000 documents available for viewing and electronic download.
18. Records of the Portfolio Management Task Force (Wapenhans Report) showed a dramatic decline in the performance of the Bank’s lending operations: over a third of the Bank projects completed were judged failures by the Bank’s staff. Morse et al. (1992), Report of the Independent Review, Resource Futures International.
19. John Mayne in his research on evaluation independence states that “Evaluation systems in development organisations can comprise at least four general types of independent practices: i) governance/oversight of the evaluation function; ii) using independent evaluators; iii) using external reviewers; and iv) involving a broad range of stakeholders in the evaluation process” (Mayne, 2011).

## References

- DFID (2014), “Rapid Review of Embedding Evaluation in UK Department for International Development,” DFID, London, <https://www.oecd.org/derec/unitedkingdom/DFID-Review-of-Embedding-Evaluation.pdf>.
- European Commission, “Better regulation: why and how”, [http://ec.europa.eu/smart-regulation/index\\_en.htm](http://ec.europa.eu/smart-regulation/index_en.htm) (All websites were accessed during the period January – May 2016).
- Expert Group for Aid Studies, <http://eba.se/en/> (All websites were accessed during the period January – May 2016).
- Heider, C. (2016), “Forum: Is There a Trade-Off between Accountability and Learning in Evaluation,” *Evaluation Connections*, February, pp. 13-14, [http://www.europeanevaluation.org/sites/default/files/ees\\_newsletter/ees-newsletter-2016-01-february-r10.pdf](http://www.europeanevaluation.org/sites/default/files/ees_newsletter/ees-newsletter-2016-01-february-r10.pdf).
- Mayne, J. (2011), “Independence in Evaluation and the Role of Culture”, *Inteval Evaluation Notes*, International Evaluation Research Groupe, [http://www.academia.edu/6468776/Independence\\_in\\_Evaluation\\_and\\_the\\_Role\\_of\\_Culture](http://www.academia.edu/6468776/Independence_in_Evaluation_and_the_Role_of_Culture).
- Morse, B. and T. R. Berger (1992), *Sardar Sarovar: Report of the Independent Review*, Resource Futures International, Ottawa, <http://ielrc.org/Content/c9202.pdf>.
- OECD (2010), *Evaluating Development Co-operation*, OECD, Paris, <https://www.oecd.org/development/evaluation/49898953.pdf>.
- OECD (2006), “Evaluation Systems and Use, a Working Tool for Peer Reviews and Assessments”, DAC Evaluation Network, OECD, Paris, <https://www.oecd.org/dac/evaluation/dcdndep/35857765.pdf>.
- OECD (1991), *DAC Principles for Evaluation of Development Assistance*, OECD, Paris. <http://www.oecd.org/development/evaluation/2755284.pdf>.
- Picciotto, R. (2013), “Evaluation Independence in Organisations”, *Journal of Multi-Disciplinary Evaluation*, Volume 9, issue 20, pp. 18-33, [http://journals.sfu.ca/jmde/index.php/jmde\\_1/article/view/373](http://journals.sfu.ca/jmde/index.php/jmde_1/article/view/373).
- Weiss, H., S. Menon and A. Bougrov (2015), *External review of the independent evaluation group of the World Bank Group: Report to CODE from the independent panel*, World Bank Group, Washington D.C, <http://documents.worldbank.org/curated/en/752241468185384036/pdf/99101-BR-REVISED-Box393188B-PUBLIC-disclosed-8-24-15-R2015-0157-1.pdf>.





**From:**  
**Evaluation Systems in Development Co-operation**  
2016 Review

**Access the complete publication at:**  
<https://doi.org/10.1787/9789264262065-en>

**Please cite this chapter as:**

OECD (2016), "Evaluation systems and governance", in *Evaluation Systems in Development Co-operation: 2016 Review*, OECD Publishing, Paris.

DOI: <https://doi.org/10.1787/9789264262065-5-en>

This work is published under the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of OECD member countries.

This document and any map included herein are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

You can copy, download or print OECD content for your own use, and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgment of OECD as source and copyright owner is given. All requests for public or commercial use and translation rights should be submitted to [rights@oecd.org](mailto:rights@oecd.org). Requests for permission to photocopy portions of this material for public or commercial use shall be addressed directly to the Copyright Clearance Center (CCC) at [info@copyright.com](mailto:info@copyright.com) or the Centre français d'exploitation du droit de copie (CFC) at [contact@cfcopies.com](mailto:contact@cfcopies.com).