Rapid Review of Embedding Evaluation in UK Department for International Development

Final Report

February 2014

Acknowledgements

Abbreviations

3ie	International Initiative for Impact Evaluation
AR	Annual Review
ARD	Africa Regional Department
ARIES	Activities Reporting and Information Electronic System
AsCOT	Asia, Caribbean and Overseas Territories Division
BC	Business Case
CHAI	Clinton Health Access Initiative
CLEAR	Centres for Learning on Evaluation and Results
CO	Country Office
CoE	Centre of Excellence
DAC	Development Assistance Committee
DANIDA	Danish International Development Agency
DECC	Department of Energy and Climate Change
Defra	Department for Environment, Food and Rural Affairs
DFID	Department for International Development
DG	Director General
EA	Evaluation Adviser
EES	European Evaluation Society
EESG	Evaluation and Evidence Steering Group
EiA	Evidence into Action
EMC	Executive Management Committee
ESG	Evaluation Support Group
EvD	Evaluation Department
FTE	Full Time Equivalent
GASPES	Governance and Security Pillar Evaluation Strategy
GEFA	Global Evaluation Framework Agreement
HoP	Head of Profession
IACDI	Independent Advisory Committee on Aid Impact
IC	Investment Committee
ICAI	Independent Commission for Aid Impact
ICF	International Climate Fund
IE	Impact Evaluation
J-PAL	Abdul Latif Jameel Poverty Action Lab
KII	Key Informant Interview
MB	Management Board

MCC	Millennium Challenge Corporation
MDGs	Millennium Development Goals
MENAD	Middle East and North Africa Department
M&E	Monitoring and Evaluation
OECD	Organisation for Economic Co-operation and Development
OJEU	Official Journal of the European Union
OP	Operational Plan
PDC	Professional Development Conference
QA	Quality Assurance
RAC	Reference and Advisory Committee
RAG	Red / Amber / Green
RCT	Randomised Control Trial
RED	Research and Evidence Division
REED	Rapid Review of Embedding Evaluation in DFID
SEQAS	Specialist Evaluation and Quality Assurance Service
SIDA	Swedish International Development Cooperation Agency
SIEF	Strategic Impact Evaluation Fund
SoS	Secretary of State
TA	Technical Assistance
ToC	Theory of Change
ToR	Terms of Reference
UKES	United Kingdom Evaluation Society
WAD	Western Asia Department
WMC	Western Asia, Middle East and Caucasus Division

Executive Summary

Purpose of the rapid review

Since 2009/10, there has been a drive within the Department for International Development (DFID) to strengthen the evidence base upon which policy and programme decisions are made. Evaluation plays a central role in this and DFID has introduced a step change to embed evaluation more firmly within its programmes. The primary purpose of this rapid review is to inform DFID and the international development evaluation community of the progress made and the challenges and opportunities encountered in embedding evaluation across the organisation.

The embedding vision and the embedding process

The direct impetus for embedding evaluation came with the change of UK government in 2010. The incoming Secretary of State (SoS) for International Development took steps to set up an independent body, the Independent Commission for Aid Impact (ICAI), with a strategic aim to "provide independent scrutiny of UK aid spending, to promote the delivery of value for money for British taxpayers and to maximise the impact of aid". The role of the existing Evaluation Department (EvD) within DFID was redefined, as was the role of evaluation within the organisation.

A vision was set out for DFID to:
- become world class in using evidence to drive value for money and impact and influence other donors to do the same
- drive programme design through rigorous evaluation of what works, allowing DFID to test, innovate and scale up
- take measured risks using high quality evidence of impact on poverty
- help partners to generate and use evidence (DFID 2010b).

Embedding evaluation would be achieved by a systematic change process. Key elements of this were:
- more staff in operational and policy divisions able to undertake specialist work on commissioning and using evaluations
- much clearer standards on evidence when new programmes were commissioned
- a rapid increase in work on rigorous Impact Evaluations (IEs) within programmes
- development of a new professional cadre of accredited evaluation specialists from across the organisation.

In practice, the embedding evaluation approach has fundamentally changed the way evaluation is managed and used across the organisation. Evaluation is embedded in Business Cases (BCs) for all new initiatives, and programme teams and embedded Evaluation Advisers (EAs) have become the front line for commissioning evaluations.

EvD's location, role and function shifted to guide and support evaluation activity across DFID.

Capacity to evaluate

There has been a strong drive to recruit, accredit and train staff in evaluation in DFID since 2011. There have been 25 advisers working in a solely or shared evaluation role, a further 12 advisers in roles with an evaluation component, 150 staff accredited in evaluation and 700 people receiving basic training. This is a major achievement in a relatively short period of time, and illustrates the commitment from senior management and staff across the organisation to build evaluation into DFID's policy and practice.

While the scaling up of capacity has been rapid, the depth of this capacity is less than required. The number of embedded advisory posts created is significantly fewer than envisaged at the outset, with eight of 25 advisers working 50% or less on evaluation. The distribution of these posts has also been uneven, with some Country Offices (COs) and operational departments having little or no evaluation advisory support. In several cases, EvD staff are effectively filling the evaluation roles of other departments. In some cases there has been a deliberate strategy to utilise support from other accredited staff members in a CO instead of having dedicated evaluation support. There are ten offices where one EA has to support between five and 20 evaluations using only 30-50% of their time.

The establishment of an Evaluation Cadre as a resource and community of practice for those working on or supporting evaluation has been viewed positively. The accreditation process has generated huge interest. However, concerns have been raised as to whether the cadre can support the professionalisation of such a large number to a level where they can competently commission and manage evaluations, or whether, at this stage in the embedding process, the focus should be on a smaller core group. Of those accredited to date, 81% are at the foundation or competent level, neither of which qualifies them to manage a substantial evaluation without specialist support. The link between accreditation to a cadre and recruitment into evaluation posts has not been consistently maintained, leading to concerns that some decentralised staff may not be receiving the appropriate information and support required to be effective in their roles.

DFID's focus on supporting external partners has been quite progressive, particularly in IEs. However, it has not sought to build capacity through the evaluations it is financing. Most contracting is with northern companies, which have few or no requirements to work with partners in the global south. This is a lost opportunity.

Effect on quantity, coverage and type

The embedding evaluation approach has contributed to a significant, but uneven, increase in the quantity of evaluations commissioned by DFID. These have increased from around 12 per year, prior to 2011, to an estimated 40 completed evaluations in 2013/14. Health, poverty and education areas are well covered by evaluation, while

wealth creation and governance are less so. This is a concern due to the increased focus which DFID has on these latter areas and the current shortage of evidence to support policy and programme decisions.

In terms of planning, the Africa divisions intend to evaluate a larger proportion of their programmes, both by number and value, than other divisions. The Western Asia Department (WAD), Middle East and North Africa Department (MENAD) and to a lesser degree Asia, Caribbean & Overseas Territories (AsCOT) do not appear to have embraced the need for evaluation to the same extent as the Africa and policy divisions. Given DFID's increasing focus on fragile and conflict affected states there is a need to review the extent to which programmes in these areas are currently being evaluated.

The focus of evaluation has changed to become almost exclusively programme oriented. There are very few thematic or country level evaluations planned whereas previously these types of evaluations accounted for the majority of DFID's evaluation portfolio. This presents a challenge to DFID as it seeks to synthesise the learning from individual projects and programmes into broader lessons for policy and programme planning and design.

Effect on quality

Evidence from independent Quality Assurance (QA) between 2012 and 2013 indicates variation in the level of quality by region and stage in the evaluation cycle. Pass rates (green and amber ratings on evaluation QA) indicate that Asia has the lowest rate at 54%, Africa 75%, MENAD 78%, and global programmes 77%. This variation points to potential gaps in technical support and oversight, which are also highlighted by differences in the coverage of country and regional EAs. Quality issues cut across most parts of the evaluation cycle, although recently the lack of attention paid to managing evaluation implementation to ensure quality has been highlighted as arguably the most critical challenge. Clarifications and improvements in guidance are also required and are being addressed.

Effect on demand and use

The embedding process has placed evaluation firmly within the programme cycle and thus increased the actual and potential demand. This has taken place in a context of wider organisational focus on using evidence, demonstrating results and increasing accountability for resources used. However, the institutionalisation process does appear to be drawing attention away from external audiences – stakeholders who can influence design and quality, and who, ultimately, use the evaluation findings.

There is a greater appreciation of evaluation in DFID and evaluation findings are already being used to some degree in decision making. However, the quality of evaluation management responses is variable and a clearer process is needed to ensure that these are timely and used consistently. Furthermore, there is not yet an agreed approach to disseminating evaluation findings and promoting evaluation use.

There is scope for DFID to adopt a much more systematic and creative approach to promoting evaluation use in the context of the wider evidence agenda. There are institutional routes for this, for example through the work of the Research and Evidence Division (RED) on evidence uptake, resource allocation rounds, policy processes and advisory cadres. Although some of these channels are being utilised, they have not yet been systematically identified and pursued.

In line with DFID's transparency agenda, all evaluations are now published, although visibility is limited as there is no platform on the DFID website where they can be easily accessed. There are some good examples of sharing and uptake of evaluation findings beyond immediate stakeholders, but DFID is not yet positioned to make best use of the emerging body of evaluation evidence either internally or externally.

Effect on value for money

The embedding evaluation approach has been accompanied by a significant increase in the number of evaluations which has, in turn, led to an increase in the total amount spent on evaluation. However, the average total cost per evaluation has changed little since 2010.

Externally procured evaluation costs appear to be in line with those of other donors. However, forecasts of future spending on evaluation indicate a likely increase in the median amount that DFID pays directly for evaluations. For non-impact evaluations the median budget is £200,000 and for IEs the median budget is £500,000. This represents a significant under-estimation of evaluation costs.

Evaluation accounts for a median of 1.9% of programme value, which is in line with expectations. The amount DFID spends on IEs is higher at 2.6% of programme value but this is consistent with the figures of other donors such as the Millennium Challenge Corporation and the World Bank.

Fitness for purpose

The extent to which the structures and processes are appropriate to achieve the aims of embedding evaluation within the organisation was reviewed at two levels: i) COs and other spending units, and ii) EvD.

In COs and UK based operational and policy departments, the decision to evaluate is being made predominantly within programme teams during the preparation of BCs. This has led to an imbalance across portfolios and a lack of a broader strategic focus, though the focus of evaluation within teams is of benefit to programme performance and future design.

The constraints to fitness for purpose are primarily capacity and the need to ensure the relevance and quality of evaluations. Evaluation staffing is uneven and where EAs exist they are almost always split posts with results or statistics responsibilities alongside evaluation. As the organisation focuses increasingly on improving programme management, following the recent 'End-to-End' Review of Programme

Management, it is likely that individuals in positions involving evaluation and results will face increased challenges in adequately carrying out their evaluation work. EAs will have to pay substantial attention to the utility of evaluations commissioned. In addition, a stronger recognition of the relationship between evaluation and results monitoring is required by DFID centrally. The mutuality of the dual roles at country level will need to be better understood and supported, particularly with the increased focus on programme management. This, in turn, might necessitate more central support to strategic, complex and impact evaluations, including jointly managed initiatives, to better ensure quality and utility at both the country and corporate levels.

Since being moved to RED, EvD has served two primary functions: establishing and maintaining the policy framework and systems for supporting decentralised evaluation, and providing a one-on-one technical support service to those designing and managing complex, impact and strategic evaluations. Both of these functions are appropriate in the context of a strongly decentralised evaluation function. To better support decentralised evaluation attention will need to be paid to EvD's responsiveness to demands for support, the quality of this support, and ultimately the quality of evaluation products. The Evidence Survey conducted in 2013 found that despite a positive view of evaluation, 45% of those interviewed felt that the evaluation advice and support on offer was only partially sufficient or insufficient for their needs. While the primary responsibility rests with decentralised EAs, the support from EvD is still too thinly spread. EvD needs to reconsider priorities and find ways to ensure the decentralised evaluation system is effective and efficient in delivery.

Implications and options

Although it was not in the scope of the Review to make formal recommendations, a number of suggestions arising from the findings are proposed for further consideration.

Evaluation coverage

The findings show that gaps have developed in terms of coverage, geographically, thematically and in areas such as aid modalities. These gaps are an almost inevitable result of the current situation, which includes decentralised decision making without strong coordination mechanisms. This is an issue that needs to be addressed, but without undermining decentralised ownership. It could possibly be tackled through an overall DFID evaluation strategy and through much stronger involvement of decentralised senior management in decisions to evaluate.

Support to operational units

DFID has made efforts to build the capacity of decentralised units to commission, manage and use evaluations. However, the QA of evaluation Terms of Reference (ToR), inception reports and final evaluation reports has identified gaps in current practice. In particular, there is a need to focus efforts on ensuring evaluations are improved by:

- understanding why and when to commission evaluation
- enhancing the contexts of evaluations and engaging stakeholders in an appropriate and timely manner
- selecting and implementing appropriate evaluation approaches while ensuring reliability of data and validity of analysis
- reporting and presenting information in a useful and timely manner.

These issues point to a critical need for DFID to ensure that operational units have the capacity to manage evaluations well, including engaging stakeholders during evaluations and undertaking QA of evaluation activities.

Making better use of evaluation

There is a sense that enthusiasm for evaluation has waned since the embedding initiative was started in 2010. This is in part due to staff changes, particularly at senior levels, and also because of a reduced visibility of evaluation results. To justify the amount of resources currently being spent on evaluation, it will be important for DFID's corporate centre, whether EvD or other research, evidence or policy teams, to communicate evaluation findings more broadly and at higher levels. Interest can only be maintained and increased if the use of evaluation is demonstrated.

Strengthening evaluation management

There has been considerable enthusiasm shown by programme managers for conducting IEs, which now comprise 28% of planned evaluations. EAs have expressed some concern that there may be a lack of understanding of what is required to conduct a good IE, in terms of time, resources and the technical capacity to manage such an evaluation properly. It will be important for COs to develop a more intensive engagement with EvD and research teams when they undertake these types of evaluation, and to ensure that central teams have the resources to provide this more intensive support.

For this to happen, EvD will need to make certain decisions. The current proportion of total DFID evaluations being supported is too great to provide the level of support required. To address this capacity challenge the commissioning of evaluations across the organisation needs to be managed more strategically, staffing in EvD needs to be increased, support responsibilities need to be shared with other parts of RED, or a much greater investment in staffing needs to be made at the decentralised level.

Enhancing evaluation governance and strategy

There is a need to strengthen the governance arrangements that determine DFID's central oversight of evaluation across the organisation. The current Evaluation and Evidence Strategy Group (EESG) has fulfilled its initial purpose and now requires a new mandate, more senior membership and a clearer reporting line to the Executive Management Committee (EMC) and/or the Investment Committee (IC). The mandate

needs to reflect the main purposes of evaluation in DFID – accountability for results, organisational learning and global evidence building – through generating new knowledge on development effectiveness. To do so, the mandate would include commissioning the development of a DFID evaluation strategy. This would specify priority sectors, geographies and thematic evaluation areas. The new mandate would also give the group responsibility for advising Directors General (DG) and directors of any changes required in their divisions for effective implementation of the strategy. The membership would include directors of geographic divisions and policy divisions to ensure organisational commitment and strategy implementation.

The development of DFID's evaluation strategy will address two key findings in this report. Firstly, regional and thematic coverage is variable and there is minimal involvement of relevant policy or evaluation teams in decisions to evaluate. Secondly, there is need to reinvigorate decentralised engagement with evaluation. A more strategic approach to evaluation has the potential to ensure optimal use of evaluation resources. It can thereby fulfil the overall purpose and function of evaluation as an essential tool for increasing the effectiveness and impact of development assistance.

Table of Contents

1. Introduction and Methodology

1.1 Introduction

Since 2009/10, there has been a drive within DFID to strengthen the evidence base upon which policy and programme decisions are made. Evaluation plays a central role in this and DFID has introduced a step change to embed evaluation more firmly within its programmes.

The primary purpose of this rapid review is to inform DFID, as well as the international development evaluation community, of the progress made and challenges and opportunities encountered in embedding evaluation across the organisation. The review findings will be presented to DFID's IC in early 2014 to inform decisions on DFID's future priorities and investment in evaluation.

The review:
- provides a stock-take of decentralising and scaling up investment in evaluation across a development organisation in its effort to improve the availability and use of credible evidence for decision making
- is intended to strengthen evaluation practice within and outside DFID around improved evaluation use and uptake to better inform policy and programming
- will be used to understand whether departments, COs and individuals have the evaluation support and capacity needed
- aims to guide future DFID and partner efforts to put in place the support, processes and individuals needed to ensure the relevance, quality, coverage and use of evaluations for improving policy and programmes.

The review does not provide a definitive measure of impact or value for money. Instead, it aims to provide an assessment of the effectiveness of efforts made to embed evaluation and reflect findings based on lessons learned.

Specific objectives of the review are to assess:
- what embedding evaluation has involved to date, has achieved and intends to achieve, including the ToC, assumptions, activities and resources
- the relevance of efforts made, the implementation process and coherence of the strategies and actions used to roll out the embedding approach
- the effectiveness of the approach, to date, in terms of increasing capacity (of individuals, departments and COs), and the relevance, coverage (by type of evaluation and sector), quality, demand and use of evaluations
- whether embedding evaluation is value for money, whether the learning and accountability balance is right, and whether the right capacity is in place to support achievement of intended outcomes. This relates to what needs to be strengthened, remain the same or be changed to ensure evaluation has impact within and beyond DFID, including specific findings relevant for EvD.

1.2   Delimitations of the review

The review was commissioned by DFID's EvD to focus on the evaluation function within DFID from the period 2010 to 2013. While recognising that evaluation sits within a broader institutional context, the ToR for the assignment are to document and review the evaluation embedding process alone. As a consequence, the shifts in the wider evidence agenda in DFID and the roles and performance of related functions, such as ICAI, are not within the scope of this report. The reviewers were also specifically asked not to focus on recommendations, which will come from a follow-up phase drawing from this review and other parallel pieces of work. Some suggestions and options are provided, but they are not intended to be comprehensive. The review was commissioned in October 2013 and finalised in January 2014.

1.3   Methodology

The review relies on a number of sources of evidence:
- document analysis, including minutes of meetings, programme Annual Reviews (ARs), reports and databases[1]
- semi-structured interviews with DFID staff and some non-DFID stakeholders, including those involved in the design and decision to embed evaluation
- findings from the DFID Evidence Survey (2013)
- a series of automated electronic questions posed to participants at the 2013 Evaluation Professional Development Conference (PDC)
- preliminary findings of an EvD study on the uptake and use of evaluation
- financial analysis of spending by evaluation type both before and after embedding evaluation.

In order to understand more fully the vision and objectives of the embedding evaluation approach, a ToC was developed and is presented in Chapter 2.

Further details on the methodology can be found in Annex 2, including details on the Key Informant Interviews (KIIs).

---

[1] A list of documents is given in Annex 3.

2. Embedding Strategy and Vision

2.1 Timing and context of embedding

The direct impetus for embedding evaluation arrived with the change of UK government in May 2010. The incoming SoS for International Development took steps to set up an independent body, ICAI, with the strategic aim "to provide independent scrutiny of UK aid spending, to promote the delivery of value for money for British taxpayers and to maximise the impact of aid" (DFID/ICAI 2011). This had been set out in the Conservative manifesto for the 2010 election. The establishment of ICAI was announced shortly after the change of government and it was formally launched on 12 May 2011, with a chair and three commissioners supported by a secretariat. ICAI reports directly to the International Development Committee in Parliament.

The relationship between ICAI and DFID is set out in the founding documentation, principally in terms of DFID's obligation to facilitate the operation of ICAI rather than setting out the role of evaluation within DFID. EvD then redefined its role and the role of evaluation within the organisation.

Discussions between EvD and the Independent Advisory Committee on Development Impact (IACDI) pointed to a future direction for evaluation within DFID. IACDI was established in December 2007 to provide an independent challenge function and to focus on issues of strategy, independence, quality and effectiveness for evaluation (IACDI 2007). A number of issues had been discussed in IACDI meetings, including ways to improve quality and lesson learning. IACDI commissioned a report on quality in DFID evaluation (2009). One of the findings reported was an unduly defensive attitude towards evaluation among staff. The report suggested that this should be dealt with by "changing the culture of evaluation to one in which independent high quality evaluation is championed and becomes commonplace". The report also recommended that evaluation be built into all new initiatives as they are developed (Riddell 2009).

DFID produced an Evaluation Policy in 2009, which contained a chapter on 'Developing a Culture of Learning and Evaluation'. The chapter focused on strengthening arrangements for decentralised evaluation and building collective learning and evidence based decision making. This policy contained ten action points for implementation. One was to significantly increase the number of decentralised evaluations in programmes and projects. Another was to support DFID staff to manage and develop their skills in evaluation and to increase staff skills in commissioning and management of high quality evaluations. All of these actions were taken forward in the embedding evaluation agenda.

Champions of evaluation were very important, in different ways, for pushing the embedding agenda forward. At a very high level, the new SoS for International Development (May 2010) had a key role. One informant remembers that he repeatedly mentioned evaluation in his first speech to DFID staff. Another key

champion was the new DG for Policy and Global Programmes for DFID. In his previous role as head of a CO he had fostered a culture of evaluation within the office. He carried this forward in his DG role by calling for DFID to use evidence much more in decision making and took forward the push for more and better evaluation. To assist and drive forward the embedding evaluation initiative, he formed the EESG with membership from different DFID divisions. EvD held responsibility for establishing and managing the EESG, including the secretariat function.

One informant described these two champions as providing the push factor (the SoS) and the pull factor (the DG). These champions were strongly supported by the Head of EvD. It is unlikely that this culture change would have happened so quickly, if at all, had DFID not needed to address the change in its role due to ICAI's[2] establishment.

2.2 Embedding vision and implicit strategy

EESG held its first meeting in June 2010 and agreed on the need for a vision statement. This was prepared and sets out an evaluation vision for DFID of:
- becoming world class in using evidence to drive value for money and impact and influence other donors to be the same
- driving programme design by rigorous evaluation of what works, allowing DFID to test, innovate and scale up programming
- taking measured risks using high quality evidence of impact on poverty
- helping partners to generate and use evidence.

The scoping note of July 2010 sees the key issues for implementation as addressing quality and rigour of evidence, improving self-evaluation and programme evaluability, building a culture of evaluation across the organisation, and reviewing and updating DFID's Evaluation Policy. It also notes a need for developing strong feedback loops from external evaluations to DFID's decision making and lesson learning (DFID 2010a).

Embedding evaluation would be achieved by a systematic change process which would be led by the EESG, and, in particular, by the DG for Policy and Global Programmes. Key parts of that change process were:
- more staff in operational and policy departments able to undertake specialist work on commissioning and using evaluations
- much clearer standards on evidence when new programmes were commissioned and a rapid increase in work on rigorous IEs within programmes
- initial piloting of the new approach to evidence and evaluation in seven COs across the first 12 months
- development of a new professional cadre of accredited evaluation specialists from across the organisation.

---

[2] Based on interviews with the 'architects' of embedding.

EvD's role would change from commissioning and delivering central evaluations to:

- **leading evaluation policy** for DFID, setting standards and QA of evaluations in programmes
- **supporting operational staff** in commissioning decentralised evaluations, providing professional leadership and developing rigorous IEs
- **disseminating** and sharing lessons from evaluations
- **building capacity and skill** on evaluation within DFID and among partners.

It was anticipated that the overall change process, including embedding evaluation across DFID and increasing skills and capacity for decentralised IEs, would take around two years (DFID 2010b, 2010d).

The EESG met regularly, initially every two-three weeks in 2010, and decreasing to every six weeks in 2011. A paper was presented to the Management Board (MB) in July 2010 which contained estimates for three scenarios:

- the baseline scenario, as reflected in the 2009 Evaluation Policy, projecting the number of decentralised evaluations rising from 15 (2010/11) to 73 (2014/15)
- a modest increase in evaluation work and an ambitious increase in evaluation work showing decentralised evaluations rising to 93 (2014/2015)
- the ambitious scenario, showing a rise of decentralised evaluations to 113.

The MB approved the ambitious scenario.

2.3 Theory of change

The review team developed a ToC for the embedding process, based on the vision statement as circulated to DFID staff, minutes of EESG meetings and submissions to the DFID MB over the period 2010/2011. Although there were further refinements in 2012 and 2013, as set out in the next section, the ToC represents as far as possible the intent and vision when the embedding process began. The ToC presented in Figure 2.1 links eight key areas of activity with outputs and intermediate and final outcomes. The review did not define impacts, but rather linked the ToC to the vision for embedding evaluation. It was validated by those responsible for developing the embedding agenda as a reasonable presentation of the implicit strategy behind embedding evaluation.

For further validation, those responsible for the embedding agenda were asked in interviews to identify what they thought were the main objectives of embedding evaluation. These varied, but many of the interviewees mentioned increasing ownership of evaluation by bringing it closer to the operational end of DFID. Through choosing to commission an evaluation, staff would be more likely to act on the findings.

The architects of embedding evaluation were also asked what they thought was meant by the term 'evaluation culture' which features in some of the documentation. Responses here were much more uniform and almost all respondents talked about

evaluation as an integral part of the project cycle. They viewed it as a core process rather than a separate activity, built into project design from the beginning. When asked what they thought were the most important measures taken to embed evaluation, three measures came up in most answers: the recruitment of additional EAs embedded in operational departments, the development of the professional Evaluation Cadre, and the inclusion of evaluation as an integral and funded part of the BC.

The ToC is used in the review as a tool to assess the strengths and weaknesses of the strategy.

Figure 2.1 Embedding Evaluation ToC

Goals Impact

Evaluation staff are recruited and retained

Evaluation staff have opportunity to devote time to evaluations

Spending units prioritise evaluations within programmes

Capacity exists to commission to undertake high quality evaluations

Evaluations affect views of key stakeholders

Assumption! Political support for evaluation is maintained

Vision for Embedding Evaluation:
- DFID is world class in using evidence to drive value for money and impact and influence other donors to be the same.
- Programme design is driven by rigorous evaluation of what works, allowing us to test, innovate and scale up.
- Taking measured risks using high quality evidence of impact on poverty
- Helping partners to generate and use evidence and build capability.

Restructuring:
- Move EvD from FCPD
- Embed EvD within RED
- Establish ICAI

Staffing:
- Establishment of 40+ decentralised evaluation posts
- Reduction of DFID EVD central staff from 23 to 12
- establishment of Evaluation Cadre

Building Capacity:
- Establish in-house training
- DFID Partner capacity development

Evaluation financed by programmes and incorporated into management cycles

Building global development evidence and public goods

Outreach and communications

Providing Routine Support:
- Evaluation Support Group
- External QA support
- External evaluation support

Defining Values:
- Evaluation Policy
- Evaluation Guidance
- Active learning

More staff in operational and policy divisions able to undertake specialist work on evaluations

Professional cadre of evaluation advisers

Capacity of staff in DFID is developed to manage evaluations

Evaluations are commissioned and managed by spending units

Systems and platforms for sharing lessons and evaluation evidence

Guidance notes

Evaluation policy / Evaluation strategies

Rigorous evaluation is a routine and accepted part of the policy and programme cycle

increase use of robust and appropriate evaluation designs, approaches, and methods

An expansion of decentralised evaluations where the evidence base is particularly weak / larger and innovative programmes

DFID staff and partners have the capacity to manage and conduct high quality evaluations

Clearer standards on evidence when new programmes are commissioned

DFID has a culture where staff routinely use evidence and evaluations in designing new programmes

Evaluation evidence is available and gaps are actively identified and filled

Robust evaluation is understood, demanded and used by DFID and partners

DFID evaluations uphold the highest standards in quality, independence and transparency

Activities   **Outputs**   **Outcomes**

7

3. Embedding Process

Embedding Process – Summary and Key Findings

Starting in 2010, the new UK government's commitment to maximise the impact of aid led to the establishment of the ICAI and, in parallel, the embedding of evaluation across DFID. Embedding fundamentally changed the way evaluation is managed and used across the organisation. Programme teams and embedded EAs became the front line for commissioning evaluations. EvD was restructured to guide and support evaluation activity across DFID.

3.1 The governance of evaluation in DFID

This chapter covers the governance of evaluation within DFID and the steps taken to embed evaluation over the period 2010 to 2013.

The structure for evaluation governance within DFID sits with the following committees and departments:

- **The Departmental Board** sets DFID's strategic direction, including oversight of the DFID Business Plan, advises on the implementation of DFID's strategy and policy priorities, and monitors results.
- **The EMC**[3] provides strategic direction to the management of DFID's operations, staff and financial resources and is supported by a number of sub-committees.
- **The IC** ensures that DFID investments represent good value for money for the UK taxpayers and that clear systems exist to take strategic financial decisions of the basis of evidence. This committee leads "on ensuring evaluation results are robustly taken forward and lesson learning embedded in the wider organisation"[4].
- **The EESG** is a sub-group of the IC. It was established to lead and direct the process of developing a culture of embedded evaluation.
- **RED** is responsible for making DFID more systematic in using evidence, innovation and learning through rigorous research, evaluation and active engagement with policy makers.
- **EvD** sits in RED to provide global oversight, high quality support and training for all DFID evaluation activity.

All DFID departments have an Operational Plan (OP) spanning the period 2011-15 which is updated annually. Among other things, the OP reflects details for Monitoring and Evaluation (M&E) within that department. The decision on what to evaluate rests with the operational unit but should be influenced by the key messages informed by the IC and passed down from the EMC to regional directorates and then COs. In 2010, EESG was established, as a sub-group of the IC, to lead and direct the embedding evaluation initiative within DFID.

---
[3] Preceded by the DFID MB, which had a slightly different function.
[4] Source: IC ToR, 2011.

Evaluations are planned and commissioned by country, regional and global programme teams for both bilateral and multilateral programmes. Operational units are supported in this by EvD, embedded EAs, the Procurement Group and the regional research hubs. In addition to directly procuring support for technical expertise and evaluations, DFID actively encourages the use of partners' independent evaluation arrangements. At the regional level, the Africa Directorate and MENAD established EA posts to backstop programme and country EAs.

Figure 3.1 below illustrates the structure of evaluation within DFID as it currently stands. EvD has a 'hands on' role to support the embedded EAs to implement policy and commission evaluations. To assist with commissioning evaluations, EvD established the Global Evaluation Framework Agreement (GEFA) and also arranged for the Specialist Evaluation and Quality Assurance Service (SEQAS), which is a resource to provide technical expertise on a demand led basis.

Figure 3.1 Evaluation Governance in DFID
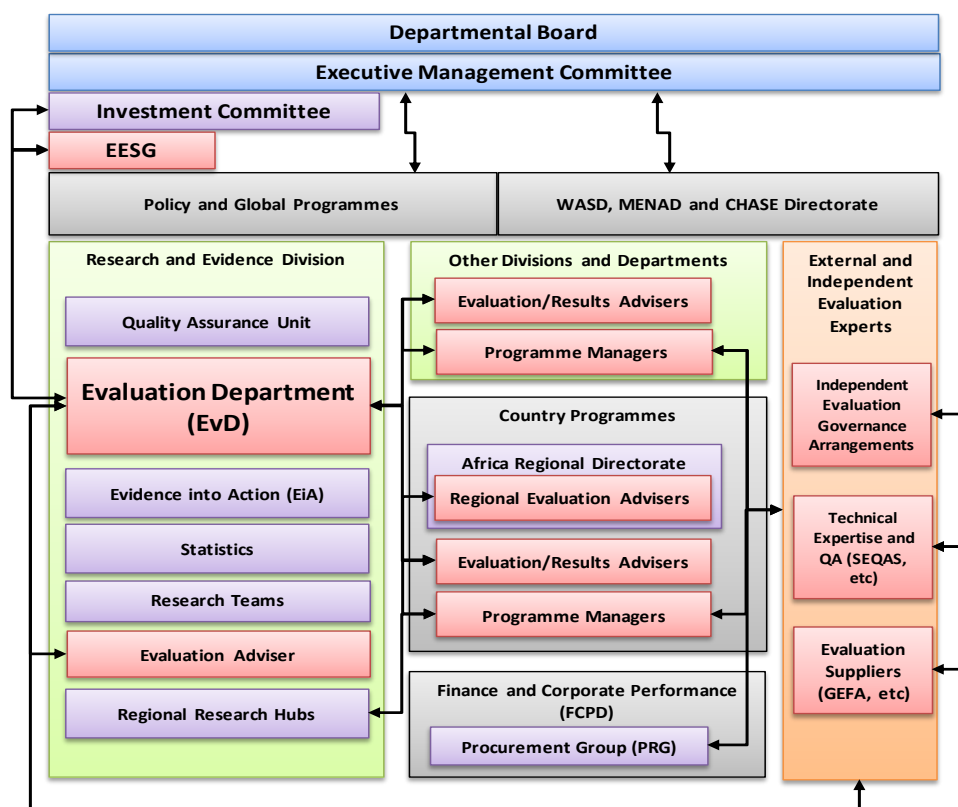
Figures 3.1 and 3.2 show how evaluation planning, management and use are interlinked across DFID teams. Similarly, the use and communication of evaluation evidence is supported by EvD and the Evidence into Action (EiA) team in RED and decentralised EAs in other departments.

Figure 3.2   Evaluation Process in DFID

- **Evaluation Vision** - Departmental Board, executive management, IC, EESG and EvD
- **Evaluation Policy** - EESG and EvD
- **Strategic Direction** - Policy and Programme teams

- **Capacity Building** - EvD and Programme teams
- **Evaluation Selection** - Programme teams
- **Evaluation Procurement and QA -** Programme teams, EAs, EvD and PrG

- **Management Responses and Decisions** - Programme managers
- **Communications** - Programme teams, EvD and EiA team

Strategy and Planning

Evaluation Management

Communications and Use

3.2. Steps taken to embed evaluation between 2010 and 2013

The headline activities of embedding evaluation are outlined in Figure 3.3. These activities were part of a systematic change process led by EESG, and guided by an implicit ToC. A more detailed version of the timeline is provided in Annex 4 to reflect other significant events that formed the wider context of embedding evaluation.

2010 was an important year for embedding evaluation. DFID set up two separate steering groups, one to work on creating ICAI and the other to lead on embedding evaluation within DFID. These groups were then merged to form EESG, chaired by the DG of Policy and Global Programmes, with a remit to oversee the process of developing a culture of evaluation and use of evidence in DFID. EESG led on engaging with DFID's MB, which took some key decisions around the scope and ambition of the agenda, and with divisional directors.

In April 2011 EvD, which until that point part reported to the DG of Finance and Corporate Performance, joined DFID's growing RED, to sit within the 'Evidence and Evaluation' pillar alongside the research and evidence teams. Strengthening evaluation was a major priority to ensure DFID and others learned from what DFID does and to put this learning and experience into use to ensure high quality spend.

Throughout 2010/11, significant achievements were made with the embedding evaluation approach, including the rapid scale-up of activity, with evaluation strategies prepared by some COs[5], external recruitment, accreditation rounds and the establishment of the Evaluation Cadre. This period also saw an increased focus on support for evaluation capacity development, including IE (Strategic Impact
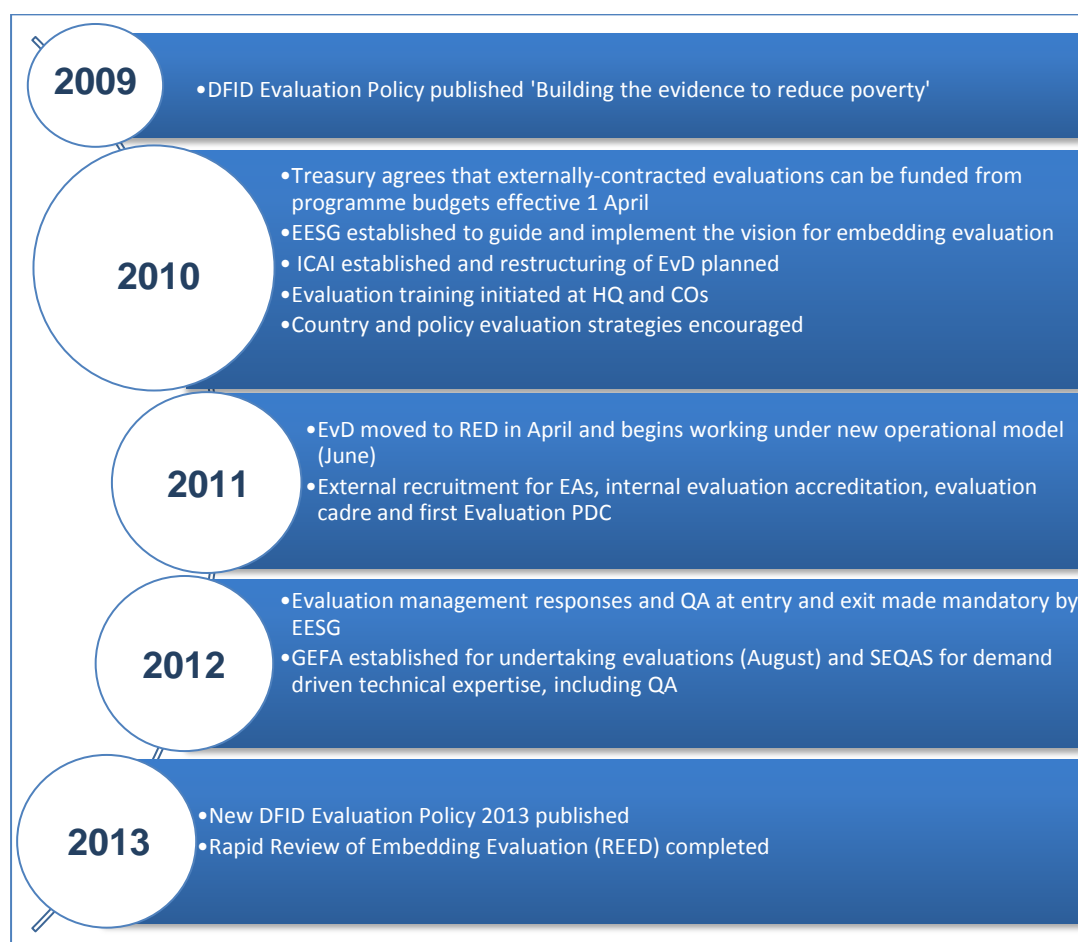
---

[5] EESG encouraged evaluation strategies but this was not mandatory.

Evaluation Fund [SIEF], International Initiative for Impact Evaluation [3ie], Abdul Latif Jameel Poverty Action Lab [J-PAL]).

In June 2012, the IC was encouraged by the increased number of planned evaluations, but also noted the challenges posed by DFID's devolved working practices and the risks to evaluation quality. Accordingly, there was an increased focus on the status of planned evaluations and the fit with DFID priorities. Shortly afterwards, EESG's role was adjusted to become an evaluation leadership group led by the Head of EvD with support from the DG of Finance and Corporate Performance. This decision was made because: a) the embedding evaluation change process was well underway, and b) the IC had increased its level of focus on evaluation activity.

The new Evaluation Policy was published in May 2013 following wide consultation with, among others, private foundations, southern partners and non-governmental organisations.

Figure 3.3 Timeline of Embedding Evaluation

2009
•DFID Evaluation Policy published 'Building the evidence to reduce poverty'

2010
•Treasury agrees that externally-contracted evaluations can be funded from programme budgets effective 1 April
•EESG established to guide and implement the vision for embedding evaluation
• ICAI established and restructuring of EvD planned
•Evaluation training initiated at HQ and COs
•Country and policy evaluation strategies encouraged

2011
•EvD moved to RED in April and begins working under new operational model (June)
•External recruitment for EAs, internal evaluation accreditation, evaluation cadre and first Evaluation PDC

2012
•Evaluation management responses and QA at entry and exit made mandatory by EESG
•GEFA established for undertaking evaluations (August) and SEQAS for demand driven technical expertise, including QA

2013
•New DFID Evaluation Policy 2013 published
•Rapid Review of Embedding Evaluation (REED) completed

3.3 Evaluation support and QA

In June 2012, EESG accepted EvD's recommendation to mandate that all evaluation products be quality assured at entry and exit stages. Prior to this decision, QA was strongly encouraged by EvD as 'best practice' and a panel of independent suppliers was available to undertake the QA function. EvD then established two external services for evaluation activity, one to assist with undertaking evaluations and the other to provide both specialist technical advice and a QA role. QA is carried out by the external provider unless, for joint evaluations, there is an equivalent function available through a partner organisation.

Global Evaluation Framework Agreement

The GEFA was established in August 2012 to give DFID staff ease of access to a panel of pre-qualified suppliers able to deliver quality evaluations which adhere to the Quality Standards for Development Evaluation of the Organisation for Economic Co-operation and Development (OECD) Development Assistance Committee (DAC). The GEFA is available for use with all DFID funded investments and also extends to programmes funded through the International Climate Fund (ICF) and jointly managed by DFID, the Department for Energy and Climate Change (DECC) and the Department for Environment, Food and Rural Affairs (Defra).

In 2012, a framework contract was awarded to 27 suppliers covering the detailed design and/or delivery of a variety of evaluation types across a range of thematic and geographic areas. For evaluations above the EU threshold, individual call-down contracts are awarded as a result of a mini-competitive process. Only suppliers who match the thematic sectors and evaluation type identified in the ToR for each individual contract are invited to tender. As at 2 December 2013, 25 contracts had been awarded with a total value of £29.8 million.

SEQAS

The SEQAS commenced in December 2012. This is a supplier consortium resource, established for three years, to provide both evaluation specialist expertise as required, and independent QA at entry and exit stages of DFID funded evaluations. As with the GEFA, the service is also available to DECC and Defra for programmes in the ICF.

The original design for the SEQAS was as an interim arrangement to assist DFID with an anticipated significant requirement for demand led specialist Technical Assistance (TA), as a result of the huge up-scaling of evaluation activity across DFID. In practice, however, the expected level of demand has not (so far) materialised. This could suggest that DFID has a good level of in-house skills or that staff are not fully aware of the services offered under the SEQAS. Currently, QA is undertaken by the SEQAS or a partner's similar governance process, if available. The introduction of mandatory QA has resulted in a high volume of products to be scrutinised and this has been the main focus of SEQAS tasks since the contract was established.

4. Quantity and Coverage

Quantity and Coverage – Summary and Key Findings

The embedding evaluation approach has contributed to a significant increase in the quantity of evaluations conducted in DFID, from around 12 per year prior to 2011 to an estimated 40 in 2013/14. The coverage of evaluations across the organisation is variable.

In terms of **thematic sectors**, those relating to the traditional Millennium Development Goals (MDGs) of health, poverty and education are well served by evaluation, while sectors such as wealth creation and governance are not so well covered by planned evaluations. This is a concern because of DFID's increased focus on these latter areas and the current shortage of evidence to support policy and programme decisions.

In terms of **geography**, the two Africa divisions intend to evaluate a larger proportion of their programmes, both by number and value, than other divisions. WAD, MENAD and to a lesser degree AsCOT do not appear to have embraced the need for evaluation to the same extent as the Africa and policy divisions. Given DFID's increasing focus on fragile and conflict affected states there is a need to review the extent to which such programmes are currently being evaluated.

The **type** of evaluation has changed to become almost exclusively programme orientated. There are very few thematic or country level evaluations planned whereas previously these types of evaluations accounted for the majority of DFID's evaluation portfolio. This presents a challenge to DFID as it seeks to synthesise the learning from individual programmes into broader lessons for policy making.

4.1 Decision to evaluate

The quantity and coverage of evaluations is driven by the process of deciding whether or not to evaluate a programme. The embedding approach has seen this decision entirely delegated to COs and in many cases to programme teams themselves. A number of key factors influence the decision to evaluate:

Country Evaluation Strategies: Deputy directors responsible for operational or policy departments have the authority to make the decision on whether or not to evaluate a programme. In some cases this decision making process is formalised in a country evaluation strategy or thematic evaluation strategy. Of DFID's 29 COs, 20 have evaluation strategies while only two thematic areas (Governance as well as Private Sector and Growth) have evaluation strategies. These strategies define the criteria for selecting programmes for evaluation and 75% of them list the actual programmes to be evaluated.

Where evaluation strategies exist, it is not clear how strategic the process is for determining what programmes are to be evaluated. Evaluation strategies are relatively new and at least one CO will undertake a review of its country strategy after a year to assess any weaknesses. The decision to evaluate appears to be taken within a CO without reference to the relevant policy team or to the relevant Head of Profession (HoP) for the sector.

EvD Guidance: EvD provides guidance on when to evaluate and recommends a number of criteria to be considered [6]. EvD does not have any authority to insist on a programme being evaluated, what approaches are used or what questions are asked.

BC requirements: corporate requirements for BCs now include a requirement to discuss evaluation plans for all projects. Projects over £40 million are further subject to more detailed QA and challenge by EAs. This has facilitated an increased awareness of the need to consider evaluation as a corporate requirement at the start of the project lifecycle.

The decision to evaluate is highly decentralised with heads of offices and their staff driving the decision making. There may be a need to rebalance the extent to which central expertise, both evaluation and policy expertise, is brought to bear on the decision making process.

4.2   Quantity of evaluations

The quantity of evaluations has increased since the embedding process began, with 25 evaluations published in 2012/13[7] and a maximum of 60 expected to be completed in 2013/14, though data from November 2013 suggests is may be closer to 40[8]. This compares to an average of 12 centrally funded evaluations per year in the period 2005-10 (Drew 2011) (see Figure 4.1). Beyond 2013/14, a further significant increase is anticipated, with 120 evaluations planned for publication in 2014/15 and a further 239 evaluations in the years to follow. This compares to the 113 evaluations in total included in the 'ambitious' 2010 proposal put to the MB in 2010. It should be noted that an unknown number of decentralised evaluations may have been published during 2005-10 and are not included in this data. Pockets of strong evaluation practice are known to have existed in some COs, for example, India. Also, a small number of high profile decentralised evaluations are referenced in certain reports (Riddell 2009) but no clear picture of quantity is available.

Despite the lack of data prior to 2010, there is general agreement among interviewees that the quantity of evaluations undertaken by DFID COs and other departments has increased substantially since embedding evaluation began in 2010.

---

[6] Criteria are i) all programmes over £5 million, ii) innovative or pilot programmes, iii) programmes for which the evidence base and ToC are weak, iv) programmes which fall within UK government/DFID policy priorities, v) the evaluability of the programme, vi) stakeholder considerations and vii) whether the intervention is contentious.
[7] Years are financial years, covering a 12-month period.
[8] In the period April-November 2013, 24 exit reports passed through the QA function. At this rate, the total number of evaluations is more likely to be closer to 40.

Figure 4.1 Number of Evaluations Published by DFID[9]
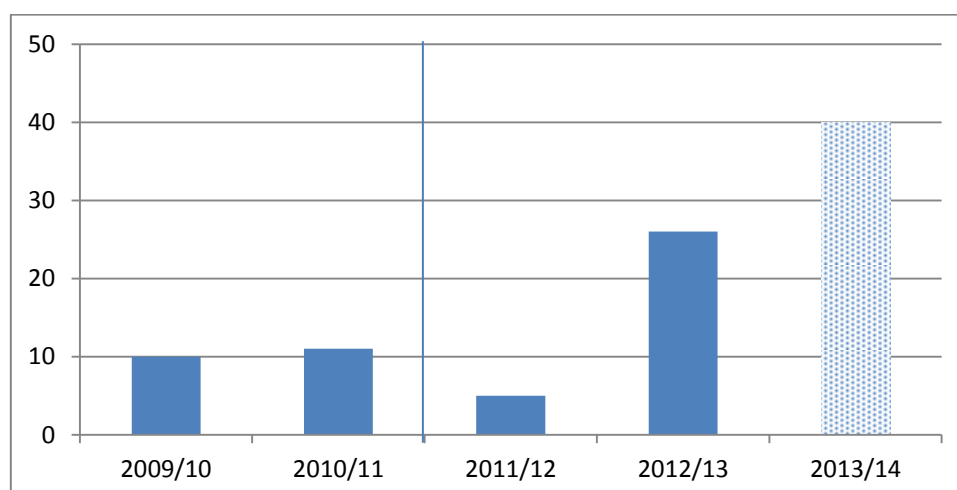
4.3 Thematic sector coverage

There are wide disparities in the forecasted evaluation coverage across different thematic sectors of DFID's portfolio. Figure 4.2 below shows that programmes relating to the MDGs of health, education and poverty/hunger/vulnerability are more likely to be evaluated than programmes in other sectors. Humanitarian, governance and wealth creation programmes are the sectors that are the least likely to be evaluated. This is a concern due to the planned increase in focus on programming in these areas, particularly wealth creation.

Figure 4.2 Thematic Sector Evaluation Coverage[10]

---
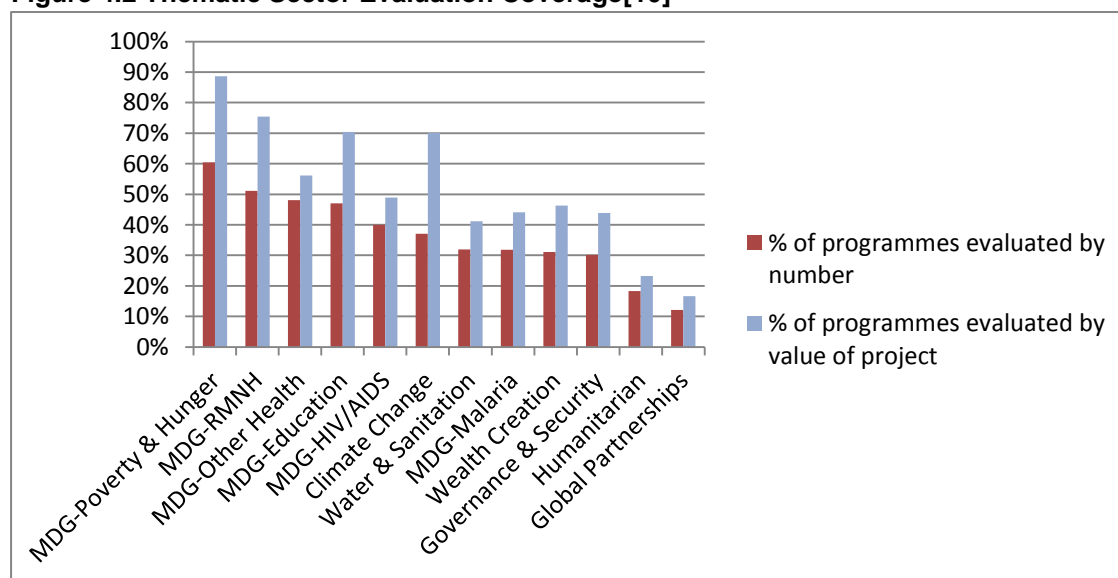
[9] Source: EvD data and EvD evaluation database. 2013/14 data is based on a forecast estimate from the 60 evaluations that are scheduled to be published. No data for decentralised evaluations prior to 2012 is available.

[10] Source: EvD and EvD evaluation database of planned evaluations.

When looked at by project value (Figure 4.2 above), the pattern of evaluation coverage by sector is similar to the pattern observed when analysed by number of projects. The exceptions are Education and Climate Change. In these sectors, higher proportions of projects are expected to be evaluated by project value than by project number, due to a number of large programmes that will be evaluated.

4.4 Geographic coverage

In terms of evaluation coverage across different geographical areas and divisions, programmes in Africa are more likely to be evaluated than programmes in all other regions and divisions. Figure 4.3 shows that the number and value of projects that are expected to be evaluated is higher in the two Africa divisions, with over 60% of projects by value and 33-45% of projects by number expected to be evaluated.

Figure 4.3 Geographic Evaluation Coverage[11]
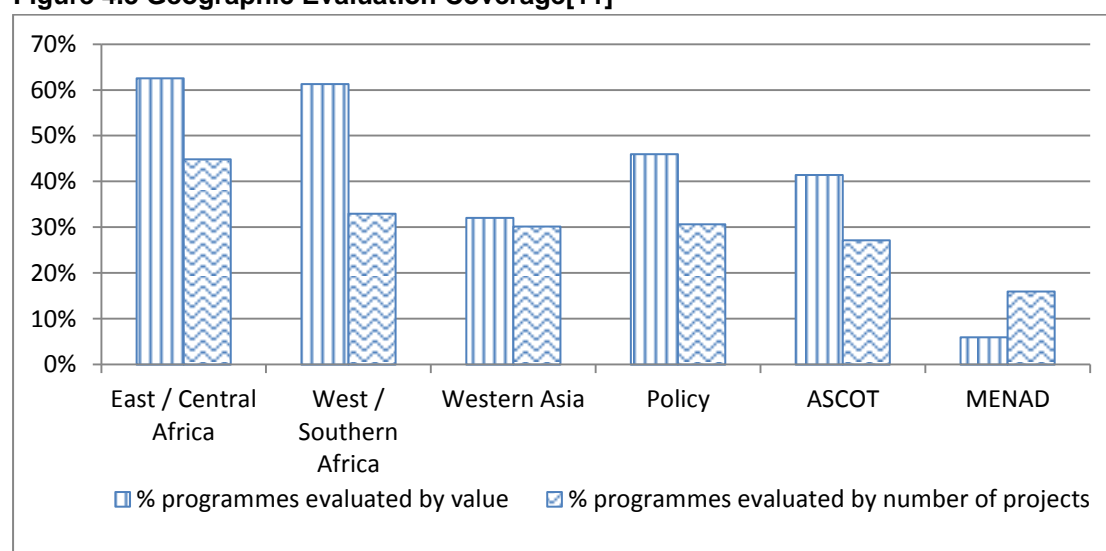
In terms of project number, most divisions expect to evaluate between 27% and 33% of their projects, with East and Central Africa evaluating more and MENAD evaluating less. In terms of project value, four divisions expect to evaluate more than 40% of their programme portfolio, with only WAD and MENAD evaluating a smaller proportion.

The expected geographic coverage is also reflected in other measures of evaluation coverage. The two Africa divisions have a higher number of EAs (Figure 4.4 below) than other divisions, and COs in Africa are more likely to have an evaluation strategy (Figure 4.5 below) than other COs. This is a result of these divisions requiring COs to have an evaluation strategy and investing in EA posts in the Africa Directorate.

---

[11] Source: EvD evaluation database of planned evaluations

Figure 4.4 Strategies[12] by Full Time Equivalent (FTE) EAs[13]
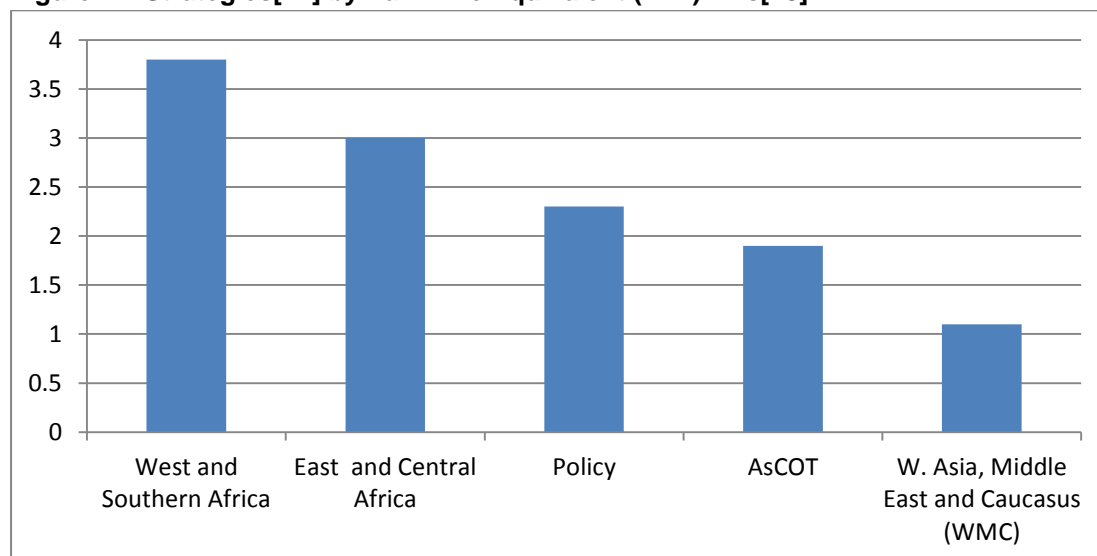
**Figure 4.5 Number of Evaluations by FTE**
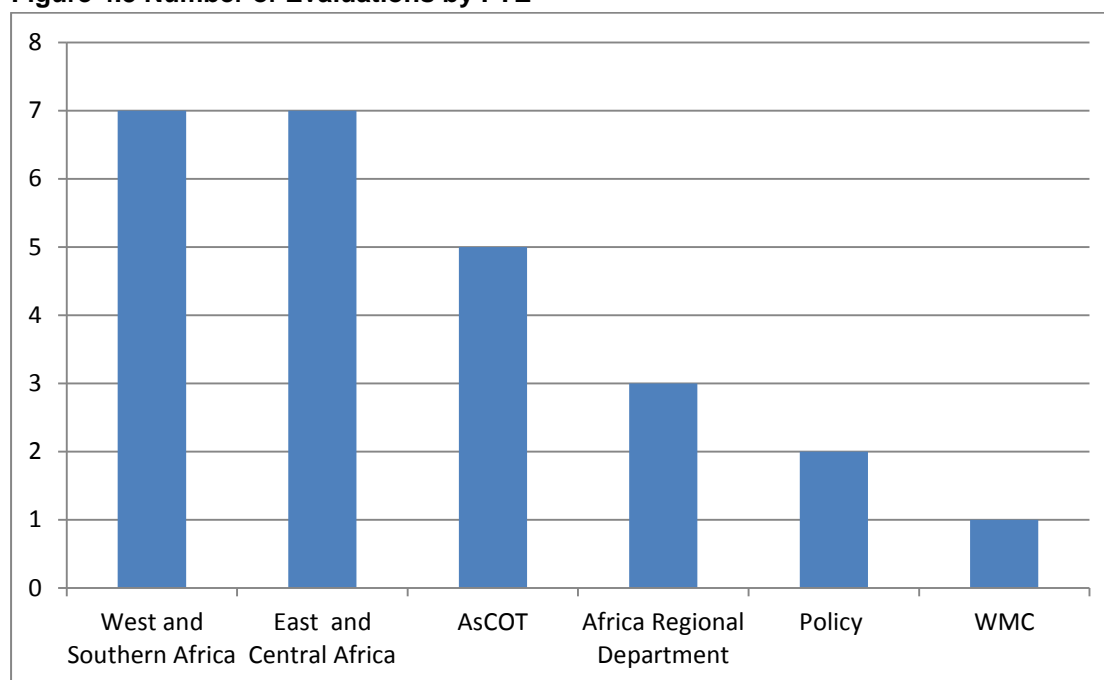
It is worth noting the change in the relative number of evaluations conducted of programmes in different regions (excluding the Policy Division) since the period 2005-10. This analysis is shown in Figure 4.6 and indicates that AsCOT and MENAD account for a smaller proportion of all evaluations undertaken in 2013 than they did prior to 2010, while all other divisions account for larger proportions, particularly the East and Central Africa Department. This data is based on the number, not the value, of evaluated projects. And as noted above, the proportion of projects evaluated by value in AsCOT is considerably higher than by number of projects. Furthermore, the changes may reflect the shift in programming spend from Asia to Africa during the

---

[12] Source: EvD analysis.
[13] Source: Evaluation of HoP. Figures include both regional department and CO advisers.

last ten years. However, the scale of the changes – in the case of AsCOT from over 35% of evaluations to less than 20% – reflects a marked decrease in the proportion of DFID evaluations taking place in this division.

Figure 4.6 Change in Regional Division Evaluations (excl. Policy Division) (Drew 2011)
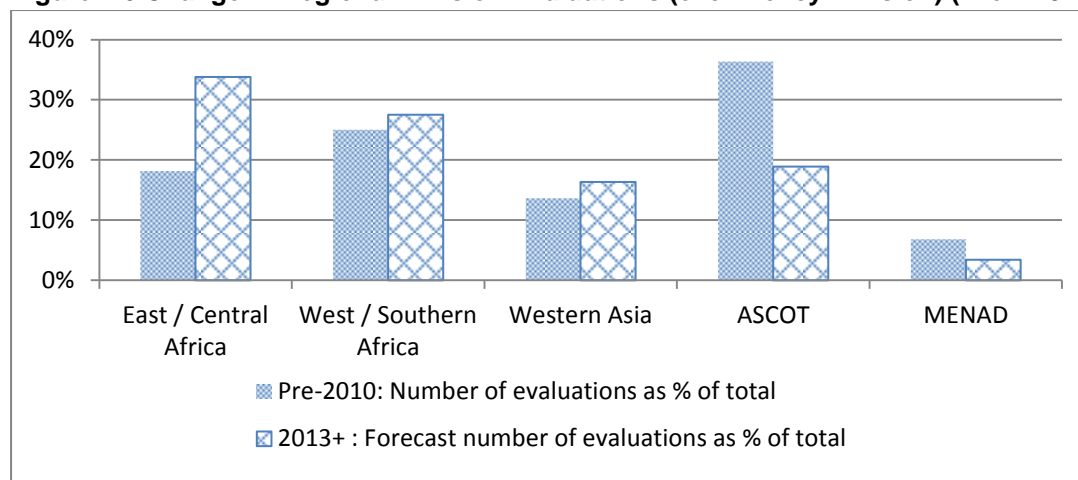
At a country level there are significant variations in programme evaluation coverage. Figure 4.7 shows the proportion of projects that will be evaluated by the top ten COs in terms of the absolute number of ongoing projects. Afghanistan and Bangladesh have a particularly low number of projects subject to evaluation compared to the total number of ongoing projects while the Africa Regional Department (ARD) and Pakistan have considerably higher than average coverage.

Figure 4.7 Evaluations in the Top Ten COs by Number of Projects[14]
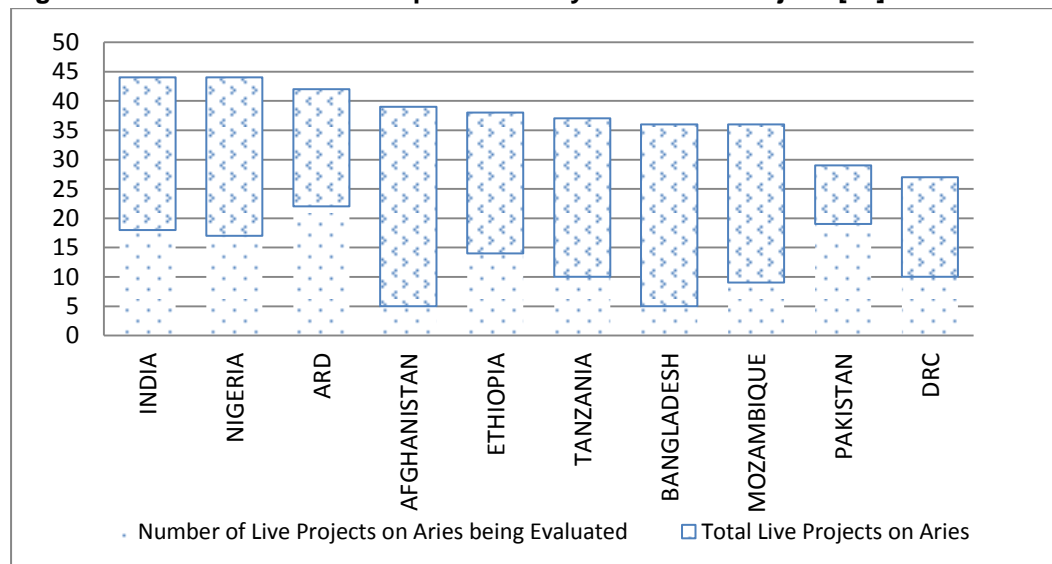
4.5   Coverage by type of evaluation

The types of evaluations undertaken by DFID have changed significantly from the 2005-10 period. During this period, central evaluations related mainly to country

---

[14] Source: EvD, ARIES, and EvD evaluation database, November 2013.

programme evaluations (46%), thematic or sector evaluations (19%) and evaluations of aid delivery mechanisms (19%) (Drew 2011).

Under embedded evaluation, evaluations are funded by programme budgets and have therefore become programme specific. There has been a decrease in the number of meso or macro-level evaluations, such as thematic sector or CO evaluations. There are currently no CO evaluations and only 7% of planned evaluations are thematic evaluations. At the same time, there has been a significant increase in the number of IEs, which are now expected to account for 28% of planned evaluations[15]. The remaining anticipated evaluation types are process evaluations (16%) or policy evaluations (4%) or are unspecified. Of the 135 categorised as IE, 80% have been verified as actual IEs, while the remaining 20% are still under verification. Of these, 70 have a completion date between 2013 and 2015. This would indicate a major increase in IEs, most of which are planned by COs.

One focus area which did feature prior to 2010 and which has reduced in number and in scope is that of aid modalities. Nonetheless there are eight budget or sector support evaluations planned, all at individual country level.

4.6   Staff perceptions of changes in coverage

Informants were asked whether they had noticed any changes in coverage of evaluations and whether they felt there were any gaps. Of the 30 internal interviewees, only ten answered this question, either because they felt they did not have enough knowledge to give an informed answer, or because of time constraints in the interview. Among those who did answer, some felt that there might be a bias towards health and education. A number of respondents felt that evaluations in governance had increased in line with spend. However, it was felt that there needed to be more synthesis studies in areas such as governance, private sector and humanitarian aid. One respondent mentioned civil society and human rights as under-evaluated areas.

Other comments focused on the type of evaluation. Half of those who answered mentioned IEs. There has been an increase in the number of IEs, but without an increase in understanding of what that entails and the alternatives to Randomised Controlled Trials (RCTs). Some respondents noted that there was no obvious central location at present for complex evaluations. These might be difficult to carry out from a CO base. External respondents commented on the absence of country programme evaluations and work on aid effectiveness.

Finally, at this year's Evaluation PDC staff were asked to vote on a series of questions, including their views on DFID's overall coverage of evaluation. Of 66 staff who answered, only 12 felt that DFID was doing the right amount overall. Five thought DFID was not doing enough and 42 felt that too many evaluations were being done and were not sufficiently strategic.

---

[15] Source: EvD, EvD evaluation database, September 2013.

4.7 External perceptions of change in DFID evaluations

A concern expressed by some staff and by members of the OECD DAC when evaluation became decentralised was that DFID's relationship with the wider evaluation community would suffer. Prior to 2010, DFID, and EvD in particular, played a leading role in the wider evaluation community. In 2010 the Head of EvD was the Chair of the OECD DAC Network of Development Evaluation (EvalNet) and DFID had played a key role in a number of joint evaluations, particularly in the lead-up to the High Level Forum on Aid Effectiveness in Accra. After embedding it was not clear who would take the lead in international engagement – DFID or ICAI – and who should be the contact point for other bilateral agencies who wished to undertake joint evaluation work.

Between 2007 and 2011, DFID participated in 15 joint thematic evaluations, and led in about one-third. These joint evaluations were a large part of the work that DFID managed. On the basis of the EvD evaluation database, it appears that of evaluations which are either completed or due to report in 2013/2014, there are three joint evaluations with other donors. Two of these are being managed by the Policy Division, on environmental issues, and one on the UN system is managed by the International Division. It would appear that this kind of evaluation has decreased substantially in number.

It should be noted that joint evaluations in general, not just with DFID involvement, have decreased in popularity. One informant commented that joint evaluations tend to focus on learning and that many agencies are now more focused on accountability. Nonetheless, DFID is still regarded as an important, professional and respected partner in the international development community. Both EvD and ICAI attend EvalNet meetings. However, ICAI does not undertake any joint work with other agencies.

There is a perception that DFID is possibly a little more introverted now. One informant feels that a decision was made that international engagement was not so important. Embedding was about strengthening evaluation internally. There is a feeling that DFID is a little less visible and its presence at the highest tables has disappeared. Nonetheless, EvD is still producing papers which are highly regarded (the recent paper by Elliot Stern on broadening the range of designs and methods for impact evaluation was cited). It has also maintained support to organisations such as 3ie and J-PAL which are producing global public goods in the area of IE.

The relationship with development partners is also strong at country level. Of 76 evaluations either completed or due to be completed in 2013/2014, just over half were commissioned by a DFID division or CO on its own. The remainder (30) were joint evaluations either with another donor (27) or with a partner government (three)[16]. Many projects are jointly funded and evaluations are planned and carried out with other donors locally.

---

[16] Calculated from EVD evaluation database.

Although it has not been covered directly in this review, it should be noted that DFID gives significant funds to a number of multilateral organisations. These organisations have also shown an increase in number of evaluations produced. Increasingly these include strategic evaluations, often in the form of comprehensive evaluations of the organisation or of the main programmes within it and linked to funding replenishment cycles. DFID's policy is to work with these organisations to improve their evaluation function, rather than evaluate the impact of its funding separately.

There has also been considerable interest globally in the embedding/ decentralisation process. EvD has had visiting delegations from Australia, Germany, Norway and South Korea interested in the process. There has also been cross-government interest in the UK.

One point made by a number of respondents was that evaluation output was less visible than in the past. One informant used the phrase "less accessible and less transparent". There appears to be a potential contradiction between the amount that DFID is investing in evaluation and the visibility of outputs.

5. Evaluation Capacity

Capacity – Summary and Key Findings

There has been a strong drive to recruit, accredit and train staff in evaluation in DFID since 2011. There have been 25 advisers working in a solely or shared evaluation role, a further 12 advisers in roles with an evaluation component, 150 staff accredited in evaluation and 700 people receiving basic training. This is a major achievement in a relatively short period of time, and illustrates the commitment from senior management and staff in EvD and across the organisation to build evaluation into DFID's practice and culture.

While the scaling up of capacity has been rapid, the depth of this capacity is less than required. The number of embedded advisory posts created is significantly fewer than envisaged at the outset, with eight of 25 advisers only working 50% or less on evaluation. The distribution of these posts has also been uneven, with some COs and operational departments having little or no evaluation advisory support. In several cases, EvD staff are effectively filling the evaluation roles of other departments. In some cases there has been a deliberate strategy to utilise support from other accredited staff members in a CO instead of having dedicated evaluation support. There are ten COs where one EA has to support between five and 20 evaluations using only 30-50% of their time.

The establishment of an Evaluation Cadre as a resource and community of practice for those working or supporting evaluation has been viewed positively. The accreditation process has generated huge interest. However, concerns have been raised as to whether the cadre is able to support the professionalisation of such a large number to a level where they can reliably design and manage evaluations, or whether at this stage in the embedding process the focus should be on a smaller core group. Of those accredited to date, 81% are at the foundation or competent level, neither of which qualifies them to manage a substantive evaluation on their own. Furthermore, the link between accreditation to a cadre and recruitment into evaluation posts has not been consistently maintained. This has led to concerns that some decentralised staff may not receive appropriate information and support required to perform their roles.

DFID's focus on external partners has been quite progressive, particularly in IE. However, it has not sought to build capacity through the evaluations it is financing, with the GEFA focused largely on northern organisations, and with few or no requirements to work with southern partners. This is a lost opportunity.

5.1   Introduction

There have been two main routes to increasing evaluation capacity in DFID. One was creating dedicated evaluation advisory posts in operational departments. The other was creating an Evaluation Cadre drawing on staff working in a wide range of roles to support the development of evaluation skills across DFID.

These two areas are closely linked, as recruitment into specialist evaluation posts draws on individuals within the cadre. Since 2010, 15 people have been externally appointed from other government departments or external organisations. The generation of internal capacity has been complemented by a programme of external capacity building work with partners and the international development community.

In assessing evaluation capacity, the quantity, spread and quality of resources have been considered, together with the effectiveness of internal and external capacity building work.

5.2   Establishment of advisory posts

The embedding evaluation process committed DFID to increasing the numbers of staff in operational and policy departments able to undertake specialist evaluation work. EAs provide professional advice to support the preparation of BCs and the commissioning, design and implementation of evaluations. They also have a QA role in ensuring that evaluations are well designed and executed.

The EESG discussed embedded evaluation posts in October 2011 and recognised that work needed to be done around identifying posts and recruitment[17]. At this time, an additional 40 posts were envisaged[18].

Table 5.1 Numbers of Advisers Working in Evaluation

Department/Division	EAs	Advisers with an evaluation component to their role
Asia and Caribbean	3	3
Africa (COs)	4.5	9
Africa (central)	3	-
MENAD	1	-
Central departments (excl. EvD)	6	-
EvD	7.5	-
Total	**25**	**12**
% Advisers in shared posts	*32%*	*100%*
FTE posts	*21*	*6*˜

˜ This estimate is at the highest end. Interviews with persons in post suggest that in some cases they spend much less than 50% of their work on evaluation.

Table 5.1 shows the number of EAs in post in DFID. Excluding the 7.5 EvD EA posts, there are 17.5 embedded posts. 46% of these (eight) are hybrid posts which include supporting results or statistics functions. There is only one full time EA post based overseas: this is in Nigeria and it has been vacant for over a year. There are a further 12 advisers who have an evaluation function as part (but not a primary part) of their role. This is significantly fewer than the 40 FTE embedded EA posts estimated to be required. The extent to which this resource is sufficient is assessed in Section 5.5.

5.3   Professionalising evaluation through the cadre

The establishment of an Evaluation Cadre in 2010 sought to recognise, embed and develop evaluation skills among DFID staff. All staff were eligible to apply for accreditation irrespective of grade, role or qualifications. Individuals can apply for accreditation at levels 1 to 4 (i.e. 'foundation', 'competent', 'skilled' and 'expert') and are assessed against five evaluation competency areas. The cadre also serves as a

---

[17] Source: Note of EESG meeting of 7 October 2010.
[18] Source: Presentation to MB, 22.7.10.

pool of highly skilled individuals to fill EA posts. However, unlike in other advisory cadres, no mandatory link was made between accreditation and recruitment. This means that, in some cases, individuals in evaluation posts are not formally accredited to the cadre.

By December 2013 four accreditation rounds had been completed. Table 5.2 shows that in total 153 individuals are now accredited to the Evaluation Cadre (this includes EAs). Of those accredited, around one-third are sector advisers. This excludes statistics advisers, who make up another 13%. EAs comprise 16% of the cadre. Generalists make up the remaining 38%, although this also includes staff in RED working in technical research roles. Some individuals have progressed through the accreditation levels, although precise figures are not available.

Table 5.2 Accreditation Figures

Level of accreditation	Numbers of individuals (Dec. 2013)
Level 1	34
Level 2	91
Level 3	22
Level 4	6

Other advisory cadres have been an important channel for promoting accreditation to the Evaluation Cadre. EvD has worked with the HoPs to encourage interested advisers to attend training and apply for accreditation. As a result of these efforts there has been increasing interest in accreditation from, for example, health and education advisers, 14 of which are now accredited to evaluation.

As part of efforts to build the cadre, DFID's first Evaluation PDC took place in 2011. It aimed to fully establish the Evaluation Cadre network, provide learning and development opportunities and discuss progress with embedding evaluation and ways to overcome challenges. Since then two further PDCs have taken place jointly with the Statistics Cadre, recognising the dual role that many evaluation and statistics advisers play and the complementarity in the work.

Accreditation to a DFID cadre recognises an individual's professional qualifications, expertise and experience. It provides the organisation with an assured standard, while enabling the individual to apply for posts requiring such expertise, with the support of a HoP[19].

5.4   Training

An important part of EvD's role is developing and delivering training to COs and other operational and policy epartments. EvD runs two courses – *Principles of Evaluation* and *Development Evaluation in Practice* – aimed at developing evaluation skills and knowledge among DFID staff.

The *Principles of Evaluation* training course is a two-day basic course aimed at introducing staff to the concepts, principles and terminology used in evaluation, and providing an overview of evaluation theory, approaches and practices. The course is aimed at staff seeking accreditation at level 1 or 2. Courses have attracted a mixture

---

[19] Insight article on advisory cadre accreditation, May 2013.

of international and locally employed staff in both programme and advisory roles. In total, 54 courses have been run between 2011 and 2013. The distribution of courses by location is shown in Figure 5.1 below [20].

These courses formed part of the 20 or so outreach engagement country visits conducted by EvD to promote the embedding evaluation approach and provide training. Internal training events are funded through the administrative budget, and are relatively inexpensive as they are delivered by staff. In total, 704 training course participants have been recorded, although this is likely to be an underestimate as participant numbers were only recorded for around half the courses. The appointment of two EAs to the Africa Directorate, whose role includes overseeing capacity building in the Africa region, probably contributed to the higher number of courses run in that region.

Development Evaluation in Practice is a four or five-day higher level course for evaluation managers and practitioners who are looking to attain level 3 skills (although the course alone does not equip individuals to be accredited to level 3). It seeks to prepare participants for commissioning, conducting and using development evaluation. This course is run in collaboration with the UK Evaluation Society and the first training was delivered in April 2012. It is also open to development partners such as consultancies and NGOs. DFID has now funded three courses and in total 38 DFID staff members have benefitted. Fewer evaluation training options are available to DFID staff within the Civil Service Learning programme, and pursuing alternative options requires a BC to apply for funding.

3iE was also contracted to deliver six workshops from 2011, with up to 30 participants in each. These were aimed at EAs and other accredited cadre members, to support the growing demand for rigorous IEs. Three courses were held in the UK and three overseas, in India, Ghana and South Africa. 3iE is now funded through the programme budget as it is providing a wider public good and is thus the most accessible higher level training option for DFID staff.

The training environment in DFID could be characterised as demand led as training is made available at the request of COs, normally through embedded EAs. In addition to CO and central training, there has been a separate track of building evaluation capacity in the professional advisory cadres and training has been delivered in cadre PDCs including Social Development and Climate Change. Some respondents noted, however, that there is scope to work more with generalist and administrative (including programme) staff who often have an important role in managing evaluations. In addition to training, there is a need to consider what further support programme staff may require in managing the increasing volume of evaluation contracts. This could play a critical role in delivering evaluations. Notwithstanding broad outreach to COs and other departments by EvD, there

---

[20] AH refers to DFID's office in Abercrombie House, East Kilbride, and PS to Palace Street (PS) to London offices.

remains a perception in some cadres that evaluation training is for EAs[21]. It is unclear whether this has affected training uptake by advisers.

Evaluation cafes and networks are considered useful as an alternative to traditional classroom forms of training, as are online training modules. However, increasingly there is a clear preference and demand from respondents for 'learning by doing' in evaluation, as this is where the real gains in expertise can be made. The pool of research and evaluation skills in DFID is not fully utilised as so much work is currently contracted out. At present, there are currently limited opportunities for the kind of hands on learning that respondents are seeking, which is potentially a valuable area to develop further.

5.5 Assessment of evaluation resources

As described above, the vision for embedding evaluation included around 40 full time embedded EA posts, but this target has not yet been met. Given the multi-pronged approach to capacity building, staffing levels should be assessed in view of the resource provided by the Evaluation Cadre as a whole. Some accredited staff can support evaluations in COs or other departments where specialist advice is limited. Figure 5.2 below combines figures for embedded evaluation posts and staff accredited at level 2 or above to show the overall potential evaluation resources by region.

Figure 5.2 Geographical Distribution of Advisers and Cadre Members at level 2 or above
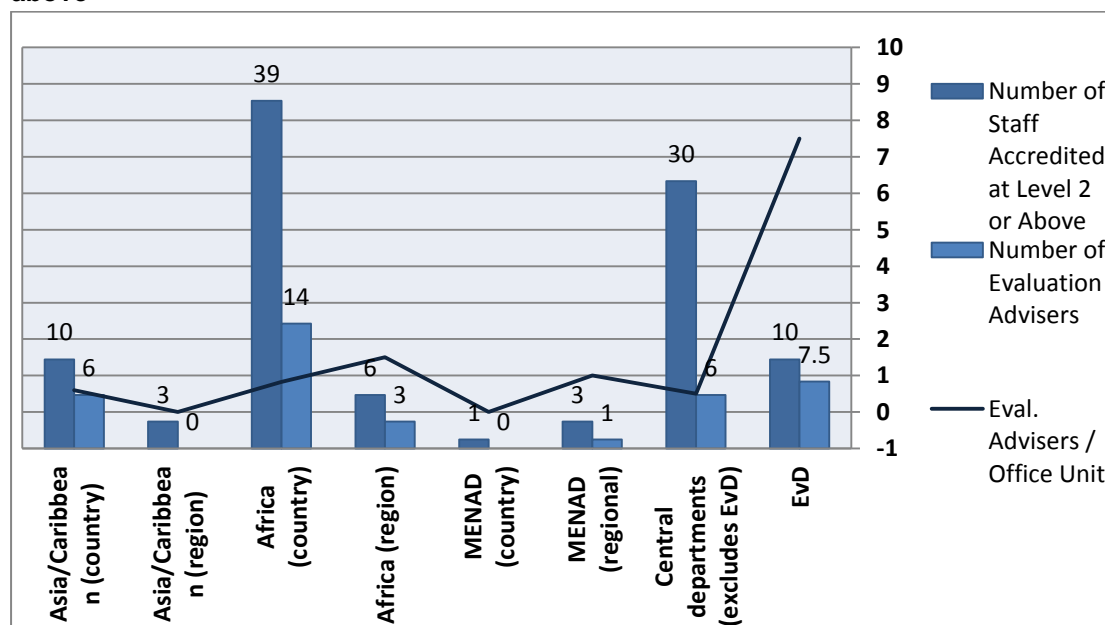
The distribution of evaluation staffing across both COs and UK based operational and policy departments is quite uneven. In the Africa Division, 14[22] out of 18 COs have some evaluation advisory resource. In addition there are two divisional EAs in

---

[21] Source: Interview with a HoP.

[22] 15 including ARD, a UK based operational department. This is distinct from the two divisional posts in the Africa Directorate.

Africa Directorate working to support evaluations and provide training to COs in the Africa region. In contrast, the Asia and Caribbean regions have six embedded EAs for 12 COs, and there is no divisional support.

Within HQ based units the distribution of evaluation resources is also varied. For example, RED, the International Financial Division, and the Conflict, Humanitarian and Security Department all have EAs. But there are none for the Human Development Department, Growth and Resilience Department, or Asia regional programmes. This variation in capacity is reflected in the data collected at the PDC.21% of respondents described capacity in their department as weak or very weak, while only 7% said it was strong or very strong.

Figure 5.2 also shows that the number of staff accredited to the Evaluation Cadre differs from region to region. This is not surprising given that country departments are of different sizes, that accredited individuals move jobs, and that some COs have encouraged accreditation more than others. Large COs typically have three-four staff accredited at level 2. Currently, only four COs and three HQ based operational or policy departments have any non-evaluation staff with level 3 accreditation. Questions remain, however, as to whether this resource is adequate, and to what extent it can compensate for dedicated evaluation advice. Accredited staff need to have time earmarked to focus on evaluation and sufficient expertise and support to further develop their skills. At least six interviewees commented that staff are unable to prioritise or make adequate time for evaluation related activities due to other pressures. Although not fully explored as part of this review, there is little indication that accredited staff work on evaluations outside of programmes which they are already managing.

In assessing resources it is important to look at the need for advice as well as supply. Figure 5.3 shows the number of evaluations planned and ongoing in COs relative to the estimated amount of FTE support assigned. This takes into accounted estimated amounts of time spent on evaluation according to the best available information.[23]

---

[23] Some individuals were interviewed by the HoP; in other cases a 50% split was assumed between results and evaluation. This may not be a complete or accurate picture of time spent on evaluation.

Figure 5.3 Relationship between Advisory Capacity and Numbers of Evaluations in COs
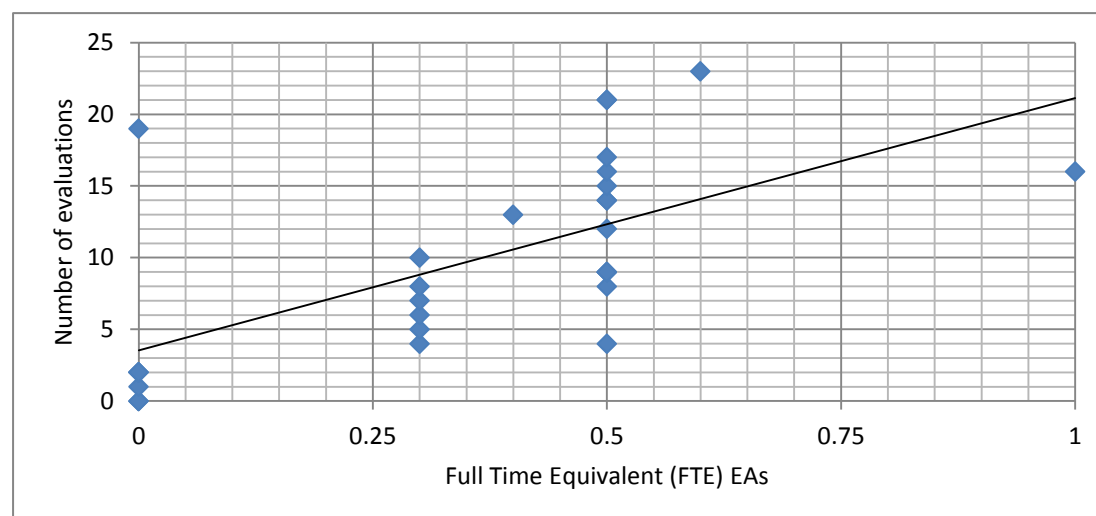
It is clear that evaluations advisory resources are thinly spread. Figure 5.3 shows five offices where five-ten evaluations are being supported with only 30% of an adviser's time, and a further five offices where 12-17 evaluations are being managed by 50% of an adviser's time. Some COs are still drawing heavily on EvD and the Africa Directorate to support evaluations, even where other in-country staff are accredited and results or statistics advisers are devoting a portion of their time to evaluation. Thus, the resources devoted to evaluation do not match the level of existing evaluation activity in DFID COs or central divisions, especially given the size and complexity of the programmes being evaluated.

This finding is supported by the interviews carried out for this review. Respondents in a range of different roles noted a significant increase in evaluation capacity in DFID. However, the majority of interview respondents said there are still not enough evaluation resources available, given the number of evaluations planned. The DFID Evidence Survey (DFID 2013b) also showed that nearly half of respondents (45%) felt that the evaluation advice on offer was insufficient or only 'somewhat sufficient' for their needs.[24]

5.6   Allocation of evaluation resources

In understanding the shortfall in evaluation capacity, it is important to understand how resources have been allocated. Evaluation posts in DFID are created at three levels: in COs or other operational departments (by heads of office/department), in regional divisions (by regional directors), and in EvD by the head of evaluation. Heads of office decide which staff to recruit depending on the priorities of the country. Directors can choose to bolster support in skill areas through creating divisional positions covering a region.

---

[24] These response categories are not ideal for reporting purposes because it is not clear how people would interpret the difference between 'Somewhat' and 'Quite'.

The high level champion for embedding evaluation, the DG of Policy and International Partnerships, communicated clear expectations of senior management which helped to support this change. Beneath this, departments have different strategies for resourcing evaluation. Some COs have put a strong emphasis on encouraging sector advisers to accredit to evaluation. In Ethiopia, accrediting staff was a deliberate strategy to generate evaluation capacity. This has enabled limited resources to be prioritised for specialist skills in other areas such as statistics, procurement and finance. Directors also have different preferences for allocating divisional level resources, and as a result evaluation is better resourced in Africa than in other regions where DFID works.

Conversely, while opening up accreditation to all staff has generated evaluation capacity in DFID, it has also led to the de-prioritisation of dedicated evaluation resources in COs, as accredited staff offer an alternative. Most EAs in COs are therefore working in hybrid posts. This has become normalised and the model has some strengths. However, a key concern is the tension in hybrid roles between working on results management and reporting where there are short-term pressures and longer-term evaluation tasks. This is exacerbated by staff turnover. PDC gathered data showed that one-third of respondents have been in post less than a year, and staff turnover was raised as a concern by some interviewees.

A further driver affecting the distribution of EA posts may be how these are funded. CO and other operational department evaluation posts can be funded through the 'front line delivery' (i.e. programme) budget as agreed with the Treasury at the last spending settlement in 2010. EvD posts are funded through DFID's administrative budget, which is more constrained than 'front line delivery', limiting the number of posts permitted.

5.7   Capacity and use of the Evaluation Cadre

Encouraging a broad skills base through a large, inclusive cadre has clearly been beneficial to embedding evaluation as it has generated interest, momentum and skills across staff in a wide range of different roles. However, the approach to accreditation has not always been systematic. A requirement for all advisory staff to be accredited to Evaluation Cadre at level 2 was introduced with the aim of making evaluation a cross-cutting skill. Implementation of this has not been systematically pursued or formally measured. In the words of one interview respondent, "it is not clear what is happening with the drive to get all advisers accredited to level 2 in the Evaluation Cadre". Respondents from within the Evaluation Cadre also noted that the potential for using 10% of advisory time to support DFID's evaluation needs and provide opportunities to develop skills has not been optimised.

Interview respondents had divergent views on the value to DFID of accrediting a large number of people. Some believed that a base level of understanding of evaluation among staff is essential for EAs to be able to work effectively with teams, and for evaluation to be fully embedded in the programme cycle. As such they thought it should not be considered a specialist activity – rather a cross-cutting

discipline that all advisers and some generalists can learn. From this perspective, having large numbers of people accredited at level 2 is valuable.

In contrast, other respondents saw little value in a large number of people reaching an 'average' standard. They advocated for the development of more in depth, specialist skills in a fewer number of individuals. In the words of one respondent, "if evaluation is a skilled area, can you expect everyone to do it"? This would mean focusing on increasing accreditation at levels 3 and 4. The two approaches need not necessarily be mutually exclusive. Other cadres employ different models, for example distinguishing between formal accreditation and affiliation to a cadre.

A further issue raised around professionalising evaluation was the Government Social Research Network. It was noted that the synergies between research and evaluation agendas in DFID could be explored now that EvD is part of RED. While research and evaluation have different purposes, there are some analytical skill sets common to people working in both areas. By addressing this, DFID could potentially attract more new recruits from other government departments.

Key questions arising are how the burst of energy around accreditation should be sustained, and how individuals who have been accredited should keep their skills up to date. Further insight into why individuals choose to become accredited, as well as how exactly this benefits their professional development and contributes to DFID's work, would be valuable. There is appetite for direction from the evaluation HoP to clarify these issues.

5.8 External partner capacity building

This section reviews the effect that the embedding agenda had on partner capacity building.

Prior to embedding

Prior to the decentralisation of the evaluation function in DFID, the support for capacity development of partners in evaluation was limited to a number of centrally managed small grants to international development associations and networks, the financing of partner country staff bursaries for evaluation training and attendance at conferences and meetings[25], and the initial support for building IE practice and supply through 3ie.

At the country level, a number of discrete initiatives in Uganda, South Africa, Ghana, Vietnam and India have been providing support to national agencies in strengthening their evaluation capacity.

---

[25] International Development Evaluation Association, European Evaluation Society, Network of Networks on Impact Evaluation and International Program for Development Evaluation Training.

Since embedding

Support for evaluation capacity development since 2011 can be gauged through an assessment of any changes in the scope and nature of direct capacity development efforts at central and decentralised levels. Indirect support can be determined by the extent to which capacity development has been an active part of the strategy to scale up evaluation across the organisation.

Direct capacity development

Within EvD and among central units there has been a considerable shift towards strengthening the evaluation capacity of partners – directly and indirectly. Since 2011, initiatives that have capacity development components totalled over £70 million, largely in the context of DFID's increasing drive for IE. The majority of these have been performing well (average latest AR score of A). Table 5.3 provides details.

Table 5.3 DFID Centrally Financed Support for International Partnership in Evaluation, Including Capacity Development

Initiative	Partner capacity building elements
Centres for Learning on Evaluation and Results (CLEAR): Initiated in 2010 – five-year support worth £2.5 million (of $30 million multi-donor World Bank managed trust fund). Support to seven centres globally serving four continents and providing regional training, TA, advisory services and networking/leadership, primarily for governments.	Progress made by CLEAR during 2012/13 is judged to be strong given the complexity and range of contexts within which CLEAR is operating, including diagnostic needs assessments, high quality training in IE and advisory work with governments on each continent.
3ie: Promotes the generation of new evidence from IEs which is relevant to policy and programme design and implementation. DFID has shown significant support for 3ie since its start up in 2008. Total allocation £32 million.	Supports capacity building in IE by providing TA to grantees who can demonstrate capability in IE but need additional support.
World Bank Gender Innovation Lab: Provides specialist support for IE of interventions that increase agricultural productivity, markets for the poor, enterprise development, jobs, skills and training, and property rights. Programmes may either (a) target women directly, or (b) be broader interventions (for example, Making Markets Work for the Poor M4P) that include economic empowerment of women as one of their outcomes. Total allocation £11.5 million.	Partnership with external researchers (northern and southern) to conduct the IEs. For data collection, the lab will partner with local research organisations – capacity building is implicit but not explicitly addressed.
World Bank SIEF: Supports IE in the areas of maternal and child health, under-nutrition, water and sanitation, and basic education. Total allocation £29 million.	Capacity building is one of the key objectives of SIEF (13% of the whole of SIEF).
World Bank Health Results Innovation Trust Fund: Supports design, implementation and M&E of results based financing mechanisms in health. Total allocation £114 million.	No information.

Initiative	Partner capacity building elements
Clinton Health Access Initiative (CHAI): Works with the Ministries of Health in Uganda and Zambia to increase demand for and capacity of health managers to solicit and use robust evidence to inform programmatic decision making. Total allocation £2 million.	Capacity building is not the primary purpose or objective of the programme but has turned out to be an important by-product of the programme.
Experiments in Governance and Politics: A network of leading scholars and practitioners working on improving research designs, standards guides and policy briefs on new evaluative research on policy relevant questions related to governance and politics. Total DFID allocation £0.185 million over two years (of total £0.74 million including Wellspring Associates and the Hewlett Foundation).	Focus on quality of research, research-practitioner interface, standards development, and a variety of mechanisms to bring actors and evidence together from various countries.
J-PAL Governance Initiative: Supports RCTs on anti-corruption and political participation. Total allocation £1.4 million.	Small element of DFID capacity building as there are two funded places on the J-PAL Executive Education Course on Evaluating Social Programmes.

At the decentralised level, 50% (ten of 20) DFID CO evaluation strategies explicitly refer to existing or planned support to partners in evaluation capacity development (Table 5.4). While in a number of cases these remain intentions that have yet to be fulfilled, this represents a considerable scaling up from the pre-embedding period. It mirrors the overall increase in focus on evaluation within country programmes, the presence of EAs, and the increase in numbers of evaluations being conducted.

Indirect capacity support

The introduction of the GEFA of pre-qualified suppliers for evaluation (in 2012) provided a further avenue for exploring the effect of the embedding evaluation initiative on strengthening partner capacity. Suppliers bidding for the GEFA were not asked to include local/national firms in their tenders, though a track record of work with partners and their inclusion in evaluation teams was requested. There is no assessment of the extent to which local suppliers have been included in specific evaluations commissioned to date under the GEFA. However, anecdotal evidence suggests that this does occur – but infrequently and often in minor roles (local contact, data collection and dissemination) rather than in lead roles, data analysis and writing.

It is clear that DFID has continued to prioritise the perceived quality of international evaluation firms and individuals over an explicit objective to build capacity through evaluation. Locally commissioned evaluations and reviews below the GEFA/Official Journal of the European Union (OJEU) procurement threshold may have a greater chance of working with local consultants. This could be due to an explicit capacity building objective, expediency or a requirement of stronger local content in the work.

As DFID has moved concertedly into commissioning rigorous IEs, this has reduced the scope for working with local evaluators, given the specialism and experienced required. This is gradually shifting as direct capacity efforts through the World Bank, 3ie, J-PAL and others increase the stock of suitably qualified national consultants in IE techniques.

Table 5.4 DFID CO Evaluation Strategies' Inclusion of Evaluation Capacity Development of Partners	
Afghanistan	Plans to build state & NGO capacity in evaluation
Bangladesh	Commitment to building evaluation capacity with partners
Ethiopia	Support to implementation of large sector based evaluations and M&E systems
Ghana	Plans to work with other donors to strengthen evaluation capacity of government, networks and agencies
India	Financial and technical assistance to the government of India on evaluation. Planned support to a new independent evaluation office
Kenya	Intention to work with other donors to strengthen the government's evaluation capabilities
Rwanda	Building of M&E capacity as a systematic priority for DFID
Somalia	Consideration of how to support new third party M&E programme
South Africa	Support to South Africa's Department of Performance, Monitoring and Evaluation
Uganda	Support to establish and finance a government evaluation facility
Zambia	Detailed plan to strengthen evaluation capacity of actors outside of DFID

6. Evaluation Quality

Evaluation Quality – Summary and Key Findings

There is no basis for comparing evaluations commissioned in the pre-embedding period with those commissioned since. Quality issues raised about previously centralised evaluations included their internal usefulness, DFID's responsiveness and technical capacity. However, centralised evaluations are different from decentralised ones in nature and focus. They may therefore have different quality related strengths and challenges.

Of those evaluations commissioned in the past 12-24 months, there appears to be variation in the level of quality depending on region, the stage of the evaluation and the quality issue concerned. When looking across regions at the achievement of green and amber ratings from external QA (i.e. those reports which 'passed'), Asia focused reports have a pass rate of 54%, Africa 75%, MENAD 78% and global programmes 77%. This variation points to potential gaps in support and oversight, which are also supported by differences in the coverage of country and regional advisers. This data will provide a baseline for determining future patterns in quality.

An assessment of current strategies and guidance shows variation between countries on issues covered and understandings. Variation is also found in the guidance available for different stages of the evaluation process. Recognition of gaps in guidance has recently led EvD to focus on improving the use of evaluation governance mechanisms.

Regarding evaluation practice, DFID Evaluation Cadre members and SEQAS quality assurers highlight a need for continued support in several areas. These include ensuring adequate consideration of stakeholders, context and the intended use of evaluations at each stage (ToR, inception, implementation, reporting, follow-up, etc.); selecting, implementing and managing evaluations in a timely manner; ensuring the validity and reliability of data sources, adequate explanations of methods used and clarity in analysis; and clearly reporting evaluation findings in a manner that is most appropriate for intended audiences and uses.
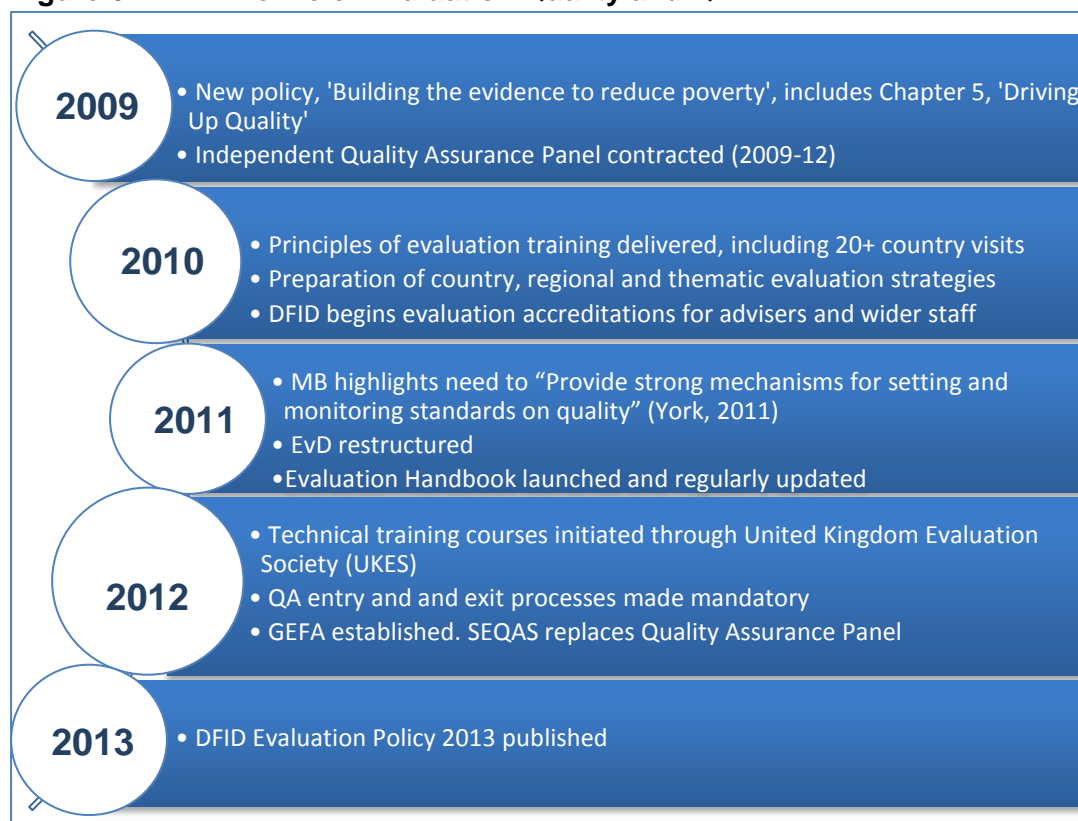
QA is perceived by many DFID staff members to be beneficial and an important part of their ongoing efforts to ensure evaluation quality. External quality assurers currently provide a solution for meeting the high level of demand that resulted from making QA mandatory in 2012. However, there are outstanding questions over whether mandating evaluations to undergo external QA is undermining partner ownership, maximising the use of EAs, and ensuring the timeliness of evaluation reporting. It may also be possible to improve quality and QA by improving governance mechanisms such as reference groups.

6.1   The drive for quality

Since as early as the 2009 Evaluation Policy, DFID has subscribed to the definition of quality agreed by the OECD DAC and captured in the DAC Quality Standards for Development Evaluation[26] (2010). In DFID, a number of steps taken to embed evaluation were directly linked to the promotion of agreed evaluation quality standards and QA (Figure 6.1).

---

[26] The DAC has set quality standards for each of the following areas: 1. Overarching considerations; 2. Purpose, planning and design; 3. Implementation and reporting; and 4. Follow-up, use and learning.

Figure 6.1   Timeline of Evaluation Quality and QA

2009
- New policy, 'Building the evidence to reduce poverty', includes Chapter 5, 'Driving Up Quality'
- Independent Quality Assurance Panel contracted (2009-12)

2010
- Principles of evaluation training delivered, including 20+ country visits
- Preparation of country, regional and thematic evaluation strategies
- DFID begins evaluation accreditations for advisers and wider staff

2011
- MB highlights need to "Provide strong mechanisms for setting and monitoring standards on quality" (York, 2011)
- EvD restructured
- Evaluation Handbook launched and regularly updated

2012
- Technical training courses initiated through United Kingdom Evaluation Society (UKES)
- QA entry and and exit processes made mandatory
- GEFA established. SEQAS replaces Quality Assurance Panel

2013
- DFID Evaluation Policy 2013 published

As shown by the timeline, embedding evaluation included training staff on the principles of evaluation, creating embedded EA posts and restructuring EvD to guide and support embedded evaluations. The need for additional QA support led to the procurement of the QA Panel (2009-12) and subsequently the SEQAS, which started in December 2012. The most recent step to guide and support evaluation quality was the publication of the DFID Evaluation Policy 2013 (DFID 2013a) (Box 6.1).

As stated by the policy, independent QA and formal management responses are mandatory, and all evaluation products should be published. Implicit in efforts made to

> **Box 6.1 DFID Evaluation Policy 2013**
>
> DFID follows the OECD DAC international quality standards for evaluation (p.1). Excerpts from the policy:
> *34. Quality pertains to personnel, process and product in evaluation.*
> *88. The aim of quality assurance is to ensure that a high standard of design, methodology, management, product quality and use of evaluation findings is achieved across DFID.*
> *89. It is mandatory for all evaluations to be independently quality assured during the design phase (entry) and at the final report (exit) stages.*
> *89. Evaluation specialists within DFID are expected to help bring the product up to, at least, minimum standards before formal quality assurance takes place.*
> *90. In addition to the quality assurance requirements, there must be a formal management response to all findings, conclusions and recommendations from each evaluation.*
> *91. All evaluation products should be made available on the DFID external website and all impact evaluations should be formally registered.*

embed evaluation is an underlying understanding that quality is linked to 'use', and that use is linked to the robustness of evaluation evidence (e.g. technical analysis and validity), responsiveness of programmes managers (e.g. management responses) and access to evidence (e.g. registration, communication, etc.).

The understanding that evaluation quality is linked to process, practice and capacity is supported by decision makers involved in the embedding process. They identified the following issues with centralised evaluation as motivations for embedding[27]:
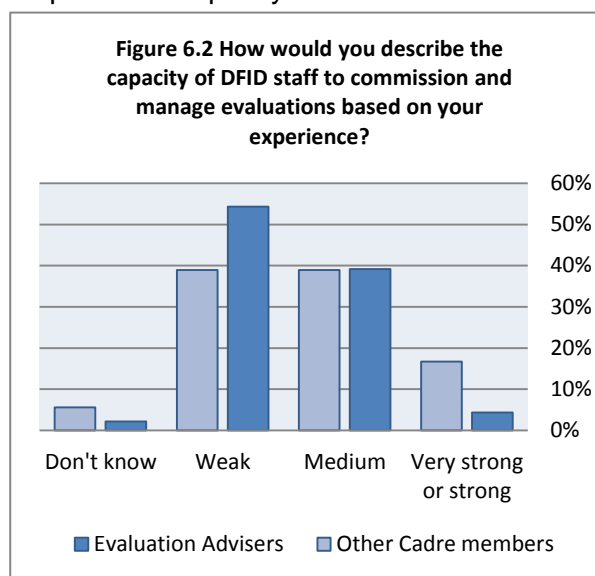
- Usefulness: centralised evaluations were perceived as under-utilised in terms of learning, programming or policy making purposes
- Responsiveness: DFID offices were perceived as needing technical support to better use evaluations and respond to evaluations
- Technical capacity: DFID was perceived as needing more human resources and technical capacity to improve evaluation capacity.

Efforts made by DFID to address these issues include capacity building (e.g. recruiting and training), guidance and support (e.g. Evaluation Handbook), and internal policies (e.g. mandatory QA). The following sections take stock of what changes are linked to efforts made, including the level of understanding of quality standards, the current quality of evaluations and the existing QA process.

6.2   Views of evaluation quality

Adhering to evaluation quality standards requires the capacity to use and understand them. DFID EAs perceive overall staff capacity as being between medium and weak[28] in this area (Figure 6.2).

As shown by Figure 6.3, the overall majority of DFID cadre members are aware of the DAC Quality Standards. However, awareness is 20% greater among individuals directly involved in evaluation related work. In terms of areas that need to be improved, the identification of users, the analysis of data, and process management were all identified by cadre members.

Figure 6.2 How would you describe the capacity of DFID staff to commission and manage evaluations based on your experience?

■ Evaluation Advisers   □ Other Cadre members

## 6.3   Current state of DFID evaluation quality

This section examines the current state of evaluation quality across DFID. The most readily available source of data regarding evaluation quality comes from SEQAS QA support. Since December 2012[29], in terms of the geographic distribution of SEQAS support requests, 22% were focused on programmes in Asia, 34% on Africa, 14% on MENAD, and 27% on multiple regions (global)[30]. Starting in March 2013, the SEQAS introduced a Red / Amber / Green (RAG) rating system, which, among other things, helps to show any regional differences that may exist in quality of products that have been quality assured.

Figure 6.3 How familiar are you with the OECD DAC Quality Standards for Development Evaluation of 2010
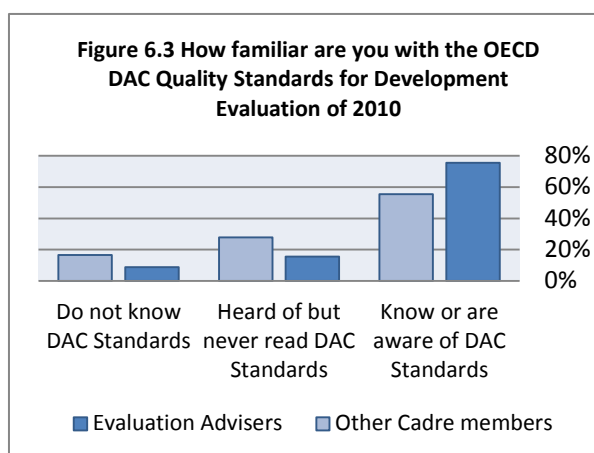
As shown in Figure 6.4, there appears to be emerging variation in quality depending on region and product types. Asia focused reports received green or amber ratings for 63% of ToR, 50% of inception reports and 25% of exit reports. Africa received green or amber ratings for 78% of ToR, 33% of inception reports and 86% of exit reports. MENAD received green or amber ratings for 67% of ToR and 100% of exit reports. Global programmes received green or amber ratings for 78% of ToR and 75% of exit reports. Keeping in perspective the very limited number of evaluations rated so far, Africa, MENAD and global programmes received green or amber ratings on at least 75% of exit reports, compared to 25% in Asia. These differences could be linked to capacity (Figure 5.2), as Asia has significantly fewer EAs per office than Africa and no divisional EAs.

Figure 6.4   SEQAS RAG Ratings from March to October 2013

---
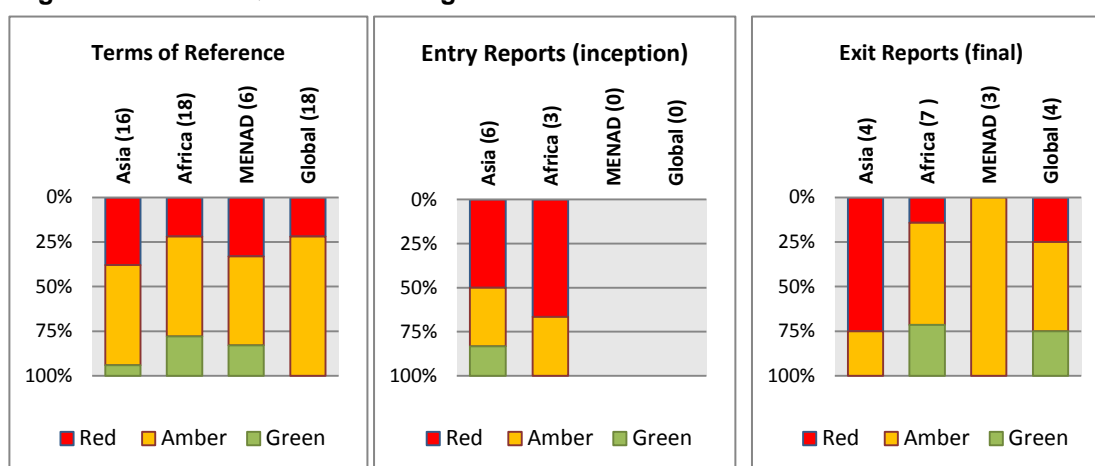
[29] SEQAS quality assured more than 132 different evaluation documents, including ToR (76 requests), inception reports (18 requests), exit reports (21 requests) and other technical support functions (17 requests).

[30] SEQAS data has not been recorded by DFID spending divisions (AsCOT, WMC, Africa, etc.) and has therefore been analysed according to regional focus categories.
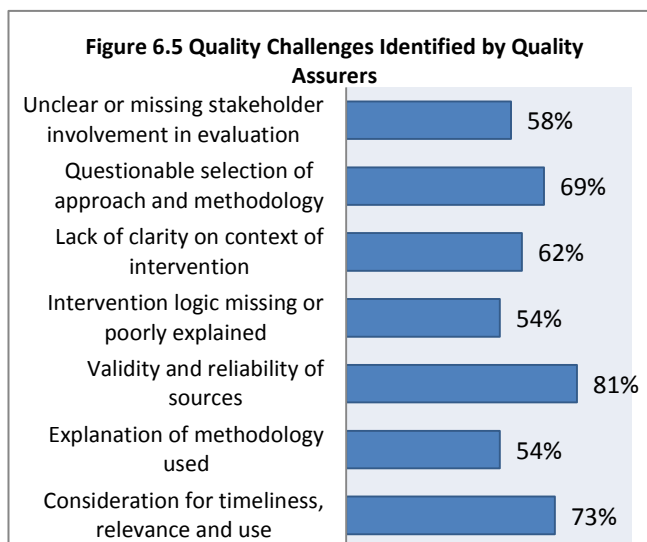
6.4 Evaluation challenges identified by external QA

Issues raised by SEQAS assurers are heavily dependent on the clarity of information presented in reports, and may be more or less relevant depending on subsequent improvements made to the documents reviewed. However, there is a significant overlap in quality issues identified by SEQAS assurers (Figure 6.5) and DFID staff. Challenges identified in SEQAS reports point specifically to the need for greater focus:

- ensuring adequate consideration of stakeholders, context and the intended use of evaluations
- selecting approaches and implementing evaluations in a timely and useful manner
- ensuring the validity and reliability of data sources, adequate explanations of methods used and clarity in analysis.

Issues of usefulness were also raised as motivating factors for embedding evaluation (Section 6.1).

6.5 Evaluation quality practice in DFID

Figure 6.5 Quality Challenges Identified by Quality Assurers

Challenge	%
Unclear or missing stakeholder involvement in evaluation	58%
Questionable selection of approach and methodology	69%
Lack of clarity on context of intervention	62%
Intervention logic missing or poorly explained	54%
Validity and reliability of sources	81%
Explanation of methodology used	54%
Consideration for timeliness, relevance and use	73%

The following is an overview of how quality is promoted and practised at each step of the evaluation process (e.g. strategy, planning, management and use).

Inclusion of quality and QA in evaluation strategies: DFID guidance states that evaluation strategies should include information about evidence gaps, evaluation stakeholders, evaluation coverage, spending and expected results (DFID 2013c). Country, regional and policy teams were encouraged to develop evaluation strategies. Offices were encouraged to seek guidance from EAs when developing strategies and asked to deposit them on the Evaluation Network Teamsite. Currently, 20 out of 29 country programmes have an evaluation strategy. There are two African regional strategies (Africa Regional and Southern Africa) and two thematic strategies (Governance as well as Private Sector and Growth).

Figure 6.6 Percentage of Evaluation Strategies Including Information on:

Category	Percentage
Evaluation Quality or Quality Standards	79%
Quality Assurance	63%
Evaluation Design	71%
Evaluation Governance	29%
Use of Evaluations	67%
External Communications	88%
DFID Partners	79%
Management Responses	41%
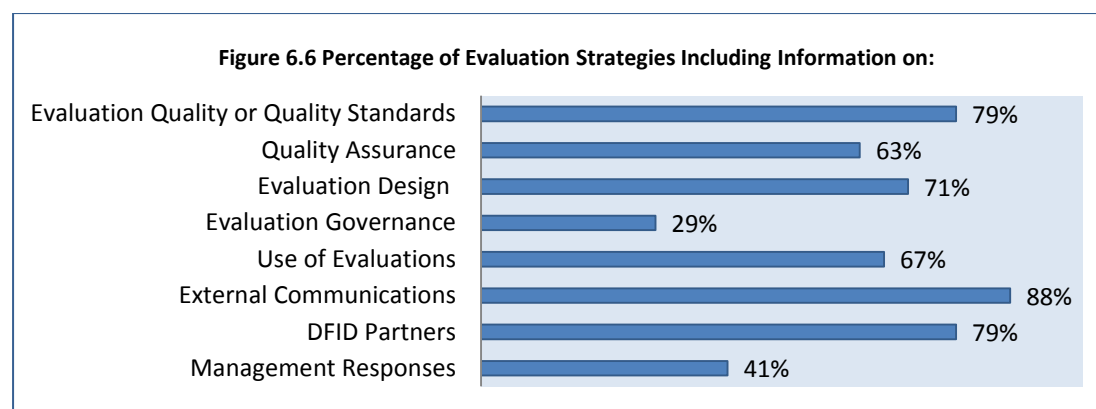
Strategies vary in length and topics covered. Figure 6.6 above provides details on the extent of coverage of quality issues in these strategies.

Although 79% of strategies mention quality, guidance varies greatly and only three strategies (Vietnam, Africa Regional, and Tanzania) make specific reference to the OECD DAC Quality Standards. The first review of an evaluation strategy covering DFID's Governance and Security pillar reveals its impact to date (Box 6.2).

> **Box 6.2 Review of the Governance and Security Pillar Evaluation Strategy (GASPES) (Grant and Chiche 2013)**
>
> GASPES was launched in October 2012 to "improve the impact of DFID's programmes and policies under the Governance and Security pillar, by ensuring our investments, decisions and project theories of change are based on the best possible evidence". Findings from a 2013 review of the strategy shows it has improved information sharing and resulted in more strategic partnerships with external actors. However, issues of leadership and clarity on the purpose of the strategy meant that GASPES was less effective at galvanising activities at centralised or country levels.

Ensuring of quality across evaluation purpose, planning and design: DFID produces evaluation guidance and training materials, including the Evaluation Handbook, how to notes and working papers, to ensure quality during the planning and design of evaluations. The Evaluation Handbook provides guidance on all stages of the evaluation process. DFID also provides support to evaluation design and ToR through support from embedded EAs, EvD, and the SEQAS. In addition, 71% of evaluation strategies provide some level of guidance on evaluation design and/or support available (Figure 6.6). The most common planning and design challenges encountered in sampled SEQAS reports are (Figure 6.5):
- the selection of evaluation approach and methodology
- reflection on stakeholder involvement
- adequate consideration for the evaluation object and scope
- need for greater clarity or attention in evaluation questions, criteria, and governance arrangements.

Evaluation implementation and reporting: Support during evaluation implementation comes from embedded EAs, EvD and external suppliers, including the SEQAS and the GEFA. Guidance on the governance arrangements for evaluations is included in 29% of evaluation strategies (Figure 6.6). The need for improved guidance and support for evaluation governance inspired a work stream initiated by EvD in 2013.

DFID staff members and SEQAS assurers both highlight the need for additional work on managing evaluation implementation. Prevalent issues in SEQAS reports are the validity and reliability of sources and explanation and/or understanding of the context of interventions. In addition, although the sample is still relatively small, 100% of the sampled exit reports cite clarity of the analysis as an issue, and 90% identify issues with the explanation of methodology and overall quality of the report. The variation in exit report quality has led EvD to begin focusing more attention on how products are published (e.g. as reviews, evaluations, or other report types).

Follow-up, use and learning: The quality of evaluations depends on the extent to which they are fit for intended purposes and uses. Guidance on the intended use and communications of evaluations is included in evaluation strategies. However, only 41% make any reference to the need for management responses. Of sampled SEQAS reports, 73% identified issues relating to the timeliness and use of the evaluation and 50% questioned plans for dissemination. In addition, Evaluation Cadre members highlight a need for DFID to focus work on the use of evaluation.

6.6   Evaluation QA

Programme staff and embedded EAs form the front line of DFID evaluation quality control. Ensuring staff members are accredited at level 2 or above is explicitly mentioned in three out of 24 evaluation strategies (Afghanistan, Nepal, and Governance). In total, there are significantly more DFID EAs and accredited staff members in Africa and central departments. However, when divided by the number of units there is currently only one EA for every two central units, five advisers for every six African offices, and two advisers for every three Asian offices. Also, no region is fully covered by either full or part-time EAs, raising questions about the availability of front line QA support. Outside of units directly engaged in programme management, the SEQAS and EvD are additional sources of QA support.

Figure 6.7 External QA Support (DFID 2012 and SEQAS 2013)
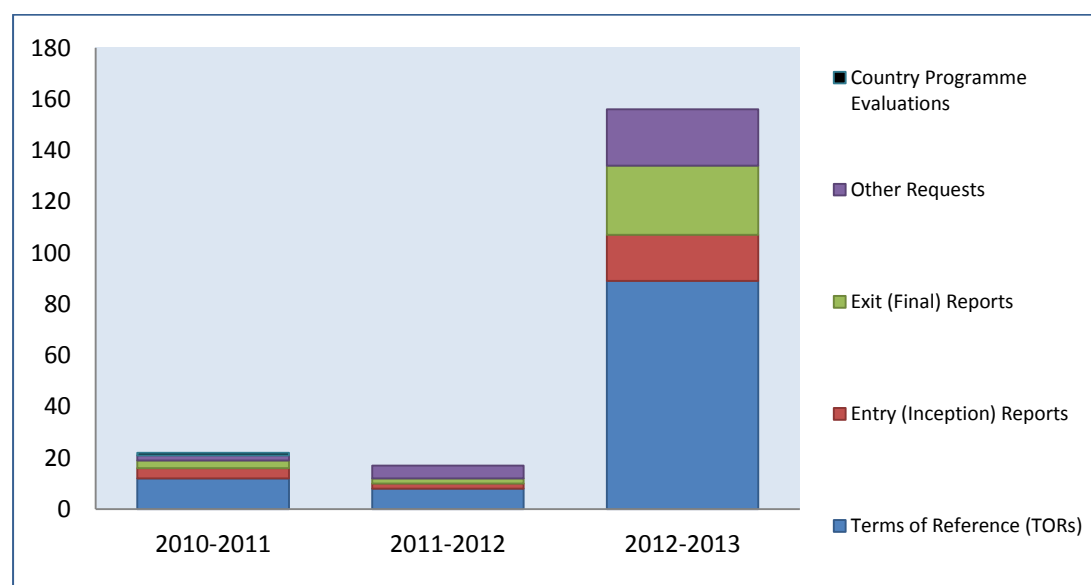
As shown in Figure 6.7, making QA mandatory in 2012 has greatly increased the number of evaluation products submitted for external QA, jumping from 17 in 2011/12 to 156 in 2012/13. Of the 156 products reviewed in 2012/13, 58% were ToR. Based on an online survey that was carried out with 17 DFID evaluation, statistics or results advisers, and 17 DFID advisers and senior managers from other professional cadres, 71% of DFID staff members surveyed found SEQAS support useful.

As reflected by DFID staff perceptions (Box 6.3), tensions are surfacing as a result of decentralising evaluation and trying to uphold quality standards. External QA options, such as the SEQAS, meet the immediate need for QA. But having external suppliers comment on the work of DFID partners is sometimes seen as undermining working relationships and ownership (although DFID filters any feedback before it is communicated to the evaluation manager).

As highlighted by interview respondents (Box 6.3), many DFID staff members value QA and credit it with improving their evaluation products (1/3 of respondents). However, several respondents questioned the usefulness of DFID's centralised QA process and its effect on partnerships. Several respondents also highlighted that there may be room for improving QA by making better use of evaluation governance mechanisms, such as reference groups and peer review mechanisms.

Box 6.3 DFID Staff Views on Evaluation QA
(based on sample of 32 staff interviews)

- Positive perceptions of current evaluation QA process:
  - mandatory QA is helpful because it removes ambiguity and prioritises quality (3 respondents)
  - QA results in improvements in evaluation products reviewed (11 respondents)
  - having an independent view is helpful when building a case for improvements (5 respondents)
  - QA early on in the evaluation process (ToR) is important and can help save money at the end (3 respondents)
- Issues identified in the current evaluation QA process:
  - QA does not ask why an evaluation is carried out in the first place
  - DFID needs to balance its use of external consultants and internal staff for implementation and QA (3 respondents)
  - the QA process is too slow and can be frustrating (3 respondents)
  - making DFID QA mandatory can undermines working relationships with partners (2 respondents)
- Amongst individuals who are not currently in evaluation focused postings (11 respondents), the majority (6 respondents) stated that QA is important for improving evaluation quality. Several (3 respondents) mentioned EvD and external QA as important resources.
- SEQAS is the most commonly referenced source of QA support (14 respondents). Several respondents saw having independent (external) support as important for supporting their case with evaluations consultants. However, the length and tone of SEQAS comments received was at times unhelpful when working with partners (4 respondents).
- EvD is recognised by many as an internal resource (7 respondents) and the same offices will often use a combination of their embedded adviser, EvD and SEQAS.
- Other QA options used include governance arrangements (external reference groups) and peer reviews (4 respondents). One respondent felt DFID needs to draw more on local expertise.

7. Evaluation Demand and Evidence Use

Demand and Use – Summary and Key Findings

Through embedding evaluation into the programme cycle and building evaluation capacity, DFID has generated demand for evaluation evidence and begun to promote its uptake and use. This has taken place in the context of a wider organisational focus on using evidence in DFID, demonstrating results and increasing accountability for resources spent. Demand for evaluation has been generated across the organisation through building evaluation into the BC, producing CO evaluation strategies, releasing programme funds for evaluation, recruiting EAs and building the Evaluation Cadre. In some instances, however, these drivers may have overtaken the structured identification of need for evidence arrived at through consultation with key stakeholders. ToC and evaluation considerations were built into DFID structures and processes during the design of the new BC format in 2011. However, the need for evaluation is still not always properly considered during programme design.

There is an appreciation of evaluation in DFID, and evaluation findings are already being used to some degree in decision making. Yet the quality of evaluation management responses is variable and a clearer process is needed to ensure that these are timely and useful. Furthermore, there is not yet an agreed set of best practices around dissemination of evaluation findings or promoting evaluation use and uptake. There is scope for DFID to adopt a much more systematic and creative approach to promoting evaluation uptake in the context of its wider evidence agenda. There are institutional routes for this, for example through RED's work on evidence uptake, resource allocation rounds, policy processes and advisory cadres. Although some of these are being utilised, they have not been systematically identified and pursued. Where evaluations produce findings later on in the programme cycle, there is a risk that momentum may be lost. Furthermore, opportunities to use evaluation evidence, including to inform ICAI studies, are limited. Keeping a focus on evaluation utility and policy relevance is vital to ensure evaluations are used to inform decisions in a changing institutional environment with many competing agendas.

In line with DFID's transparency agenda, all evaluations are now published. But there is no platform for these on the DFID website or an easily accessible platform to permit high visibility. There are some good examples of sharing and uptake of evaluation findings beyond immediate stakeholders. However, DFID is not yet positioned to make best use of the emerging body of evaluation evidence, internally or externally.

7.1 Introduction

A key element of the vision for embedding evaluation was for DFID to be "world class in using evidence to drive value for money and impact, and influence other donors to be the same". This vision built on commitments in DFID's 2009 Evaluation Policy on developing a culture of learning and evaluation. Increasing the demand and use of evaluation evidence forms part of a broader initiative, led by DFID's RED, to integrate the use of all types of evidence (e.g. research, systematic reviews) into the policy and programme cycle.

Strengthening the use of evaluation and evidence in DFID requires the promotion of an evaluation culture. Those who led the embedding evaluation drive within DFID in 2009-10 focused on integrating evaluation into existing processes and raising the profile and utility of evaluation among key users.

How has the embedding process affected the demand and use of evaluation in DFID? The ToC would indicate that, as a result of increased capacity, changes in the financing of evaluations and an increase in the number of EAs, rigorous evaluation would be a routine and accepted part of the policy and programme cycle. This implies that the demand for evaluations would increase. It also assumes that as evaluation evidence became increasingly available, it would be used in the design of new programmes and modification of existing programmes.

7.2   Evaluation demand

Evidence on the quantity of evaluations planned is presented in Chapter 4 of this report. To what extent is this growing demand for evaluation a result of changes in organisational requirements? Does it reflect a true increase in demand for evaluative evidence?

The drafting of BCs and formulation of evaluation strategies require staff to engage with the evidence base in their area of work and either commission evaluations or provide a good reason for not doing so. Country evaluation strategies point to evaluation demand which stems from local programme needs as well as wider accountability, learning and evidence building functions. These include:
- making course corrections to programmes
- scaling up or stopping existing work, or informing a second phase
- strengthening monitoring and results reporting through robust measurement
- ensuring accountability to UK government (especially for high level or novel spend)
- ensuring accountability to local stakeholders
- providing evidence on value for money
- generating evidence for wider application, i.e. global public good, especially for innovative programmes and where there are evidence gaps.

Some departments reported a greater increase in demand than others. This is influenced by factors including the level of senior management support, the presence of an embedded EA and the interest of individual sector advisers. In some instances thresholds have been set, for example to evaluate 70% of the portfolio. It is not clear what purpose such targets serve or how they affect evaluation demand.

This growing focus on evaluation is supported by findings from the DFID Evidence Survey (2013), which showed that that 87% of respondents think that evidence is important for their work. Specifically, 80% of respondents regard evaluation as at least quite important to DFID's work (DFID 2013b, p. 24). Furthermore, 64% of respondents voting at the PDC agreed that evaluation was now very well or well understood in their offices/departments. This points to an overall confidence among DFID staff around understanding the role of evaluation and its relevance to their work.

The increase in evaluation capacity across the organisation together with the release of programme funds for evaluation has undoubtedly increased demand. However, it

is difficult to untangle the links between supply of resources for evaluation (i.e. staff, funding and training) and demand to generate evidence as part of a structured identification of need. Some interviewees voiced concern that evaluation in DFID is becoming a 'tick-box' exercise in the BC which is sometimes used as lever to obtain programme approval. How the demand that has been generated is managed is critical to ensure a DFID portfolio of evaluations that will meet priority evidence needs.

The imperative to set and achieve results for the DFID Results Framework at CO level, along with wider accountability pressures to justify the growing development assistance budget, has also contributed to demand for evaluations. A key question is whether the evaluation evidence will become available in time to meet these accountability needs. A number of interview respondents highlighted that evaluations are an important source of evidence for ICAI. To be properly utilised, evaluation findings need to be available prior to the ICAI work being completed.

7.3   Management responses

It is mandatory for all DFID evaluations to have a management response and these are published together with the final report. DFID is committed to ensuring that management responses are completed, published and followed up. EvD monitors the extent to which this is happening and figures are included in reports to the IC.

In 2012/13, DFID and/or partners completed 26 evaluations of which 25 were published[31]. In 2012/13, according to the Annual Evaluation Report, only 12 of the 26 evaluations published a management response. At the time of writing, another seven had been provided to EvD, one is still outstanding, and three of the evaluations were not led by DFID and will not provide a management response. It is not however clear why there have been these delays in producing management responses.

There is variation in how much thought goes into the management responses. In some cases they were discussed by a committee of donors and implementers, while in other instances it was not clear that this had happened. Some management responses accepted all recommendations, whereas others partially or wholly rejected some recommendations. The comments provided suggest a healthy engagement with the evaluation findings. A good example is the management response to the Chars Livelihoods Programme in Bangladesh (Box 7.1).

There was also variation in the format used to write up the management response. In most cases the DFID format was used. However, this does not include space for discussion of how and when agreed recommendations will be followed up.

---

[31] One evaluation has been completed but not published due to political sensitivities.

Box 7.1 The Management Response to the Independent Impact Assessment of the Chars Livelihoods Programme, DFID Bangladesh

This was a very considered response to the impact assessment. A five-page response to the recommendations was provided, in addition to the DFID recommended format. There was also a short technical annex which addressed one major area of disagreement. The response pointed out the importance of having a lead adviser who was familiar with the latest technical and methodological aspects of this kind of evaluation. EvD and other EAs had also provided support. Stronger governance arrangements were also advised for this evaluation.

7.4   Evaluation use

The DFID Evaluation Policy highlights that "the key challenge for DFID was its commitment to promoting and improving the use of evaluations both through the process itself and in the application of evaluation findings" (DFID 2013a).

What are the conditions for evaluation uptake and use? Recommendations need to be appropriate and practical and evidence needs to be of good quality to influence policy making. People need the time and incentives to engage with the evaluation process and findings, particularly stakeholders who need to be involved from the outset. Findings need to be accessible to end-users and available at the right time. Attitudes from management about how to engage with findings can also be an enabler or a barrier. The case study illustrated in Box 7.2 is a good practice example of how evaluation findings can be used by a wide range of stakeholders throughout the evaluation process.

The DFID Evidence Survey (2013) illustrated that DFID staff recognise that evaluation is important but want more support and for findings from evaluations to be shared. Respondents also rated DFID as 'average' at learning and applying lessons compared with other organisations. This was reflected in a general finding from the focus groups that DFID does not currently make optimal use of information and lessons generated by our programmes[32]. Where evaluation findings become available later on in the programme cycle or after the programme ends, opportunities to share and utilise findings in central processes such as resource allocation or policy development are at risk of being lost.

Findings from the recent EvD uptake and use survey[33] were more encouraging. These indicate that there is already engagement among DFID programme and advisory staff – around half of respondents had used evaluations in their work "a few

---

[32] For example, pilot projects could help in generating evidence to support implementation. However, pressure to spend money means that these are rarely commissioned. Partners should be pushed to share data in meaningful ways with donors or other partners. This could include an obligation for data to be collected in a useful format.

[33] In 2013 EvD undertook a stocktake of progress on the adequacy and effectiveness of evaluation uptake and use which included a survey. The survey covered social development advisers from AH, Bangladesh, Burma, Caribbean, India, Nepal and Pakistan (most at A2 level). Surveyed programme managers ranged from B2 to A2 in grade and covered many DFID central departments and COs.

times". Uses of evaluation included to change, confirm or modify an existing programme; to understand a programme in a new way, e.g. through its strengths and weaknesses; and to influence a new programme design.

Responses from questions asked at the Evaluation PDC demonstrated that currently 11% of respondents' time is spent promoting the use of existing evaluation findings and follow-up. This is perhaps not surprising as many DFID evaluations are still in the early stages, creating an important opportunity to build in strong dissemination.

Box 7.2 Evaluation of the Community Response to HIV and AIDS, 2009-2012

This evaluation was conducted through a partnership among DFID, the World Bank and the UK Consortium on AIDS and International Development.

A phase-in approach was applied to the implementation of the evaluation as well as the sharing of findings. The validation and uptake of findings began as soon as each of the 17 studies was completed. This approach supported local ownership by working closely with stakeholders and engaging in purposeful discussion about the dissemination and application of specific findings within the local context. It generated a stream of evaluative knowledge at the national and global levels. This knowledge informed the response to HIV/AIDS well before the end of the overall evaluation exercise.

Dissemination was facilitated by country level seminars, global meetings, videos, e-communication and academic journal publications. A critical component in promoting the use of findings was engaging broader audiences through regional expert meetings, academic institutions, civil society meetings and presentations in key institutions. Products included study specific evaluation briefs for national stakeholders and decision makers, one-pagers for sharing findings with communities, and reports used for scientific dialogue.

Findings were used in different ways in different contexts to inform the response to HIV. They supported meaningful dialogue within countries, civil society and development partners. Topics included stigma (e.g. India, Lesotho), national domestic abuse (Kenya, Nigeria) and the challenges of condom distribution versus condom use (Nigeria).

The consultative process was successful because it was deliberative, purposeful and sustained over time, and the three key partners were equally committed to a robust evaluation exercise.

By Rosalia Rodriguez-Garvia, Evaluation Team Leader, World Bank (until 2013)

According to the EvD survey, the following channels are currently being used for sharing evaluation findings:
- DFID Evidence and Programme Exchange – an online evidence portal
- policy teams/forums/seminars – with internal and external participants
- advisory PDCs
- in-country sector working groups, involving partners, other donors, etc.
- cooperating partner groups, e.g. in-country statistics offices
- thematic briefs recently produced by the Evaluation Support Group
- DFID Evaluation Annual Reports (first produced in July 2013)
- published management responses (see below).

There is scope for DFID to adopt a more systematic and creative approach to promoting evaluation uptake beyond immediate stakeholders in the context of its wider evidence agenda. There are institutional structures which could be used more advantageously, for example through RED's work on evidence uptake. Although some of these channels are being utilised, they have not been systematically identified and pursued. Some may be more effective than others. For example, respondents in the DFID Evidence Survey reported that their professional networks were helpful in navigating evidence, with over 80% of respondents finding them at least somewhat helpful (DFID 2013b, p. 24). Without a proper assessment of which channels are most effective it is hard to know how best to prioritise efforts.

7.5   External publication and transparency

In line with the UK government transparency commitments in effect from May 2012, all DFID evaluation reports and related documents are published on DFID's [Development Tracker](#) and the [Gov.UK](#) external website. The DFID website should be an important communication channel for evaluation reports. However, locating evaluations on this site is difficult. Evaluations are housed in cross-government publications and international aid and development is not broken down into subtopics. The person searching has to know that an evaluation exists and search for it as browsing is not available. The Annual Evaluation Report contains a link for the series 'Evaluations completed 2012-13', but this link is not accessible from elsewhere on the DFID website as there is no evaluation platform or button.

8. Value for Money

Value for Money – Summary and Key Findings

The embedding evaluation approach has been accompanied by a significant increase in the number of evaluations, which has led to an increase in the total amount spent on evaluation. However, the average total cost per evaluation[34] has changed little since 2010

Externally procured evaluation costs appear to be in line with those of other donors. However, forecasts of future spending on evaluation indicate a likely increase in the median amount that DFID pays directly to evaluators. For non-impact evaluations the median budget is £200,000 and for IEs the median budget is £500,000.

Evaluation accounts for a median of 1.9% of programme value, which is in line with expectations. The amount DFID spends on IEs is higher at 2.6% of programme value but this is consistent with other donors such as the Millennium Challenge Corporation (MCC) and the World Bank.

8.1 Introduction

Value for money was not one of the driving forces behind the embedding evaluation approach but it provides an important context when assessing the overall effectiveness and outcomes of the embedding approach. This chapter looks at changes in the cost of evaluations and seeks to compare these costs to the best available benchmarks.

It is not possible to perform a full value for money analysis looking at the economy, efficiency and cost effectiveness of this increased investment. Cost effectiveness can only be assessed once the intended benefits of embedding evaluation have been realised. It is too early to measure the benefits such as improvements in programming based on evaluation findings. A number of indicators for economy and cost effectiveness have been derived and are estimated based on the limited available data.

The following indicators are used in this assessment[35]:
- the economy of the embedding evaluation approach, including analysis of the key cost drivers such as salaries, central projects and independent evaluation costs
- the economy and efficiency of independent evaluation costs in terms of:
    - direct evaluation contract costs
    - evaluation costs as a proportion of programme value
    - evaluation consultant daily rates.

---

[34] Total cost excluding strategic evaluation programmes.
[35] The analysis in this section is based on the available data from Aries, EvD and other donor organisations. This incomplete data set has been supplemented by assumptions and estimates where necessary and therefore can only provide an indication of the possible costs of the embedding evaluation approach. More detailed data gathering and analysis are required before action is taken based on the data included in the section.

8.2 Economy of the embedding evaluation approach

Overall cost and key cost drivers

The overall aggregate annual cost of conducting and managing the evaluation of DFID's development programmes worldwide has increased from £3.1 million in 2009/10 to just over £14.8 million in 2013/14. These figures include all evaluation related costs such as evaluation consultant fees, DFID staff salaries, capacity building and administration. During this period the number of evaluations published by DFID has increased from ten evaluations in 2009/10 to an estimated 40 evaluations completed in 2013/14.

Figure 8.1 shows the change in overall evaluation costs and the evolution of the key cost drivers for independent evaluation. Costs of central EvD salaries, administration, evaluations and minor programmes[36] have decreased from £3 million to £1.9 million during this period[37].

Figure 8.1 Total Evaluation Costs[38]
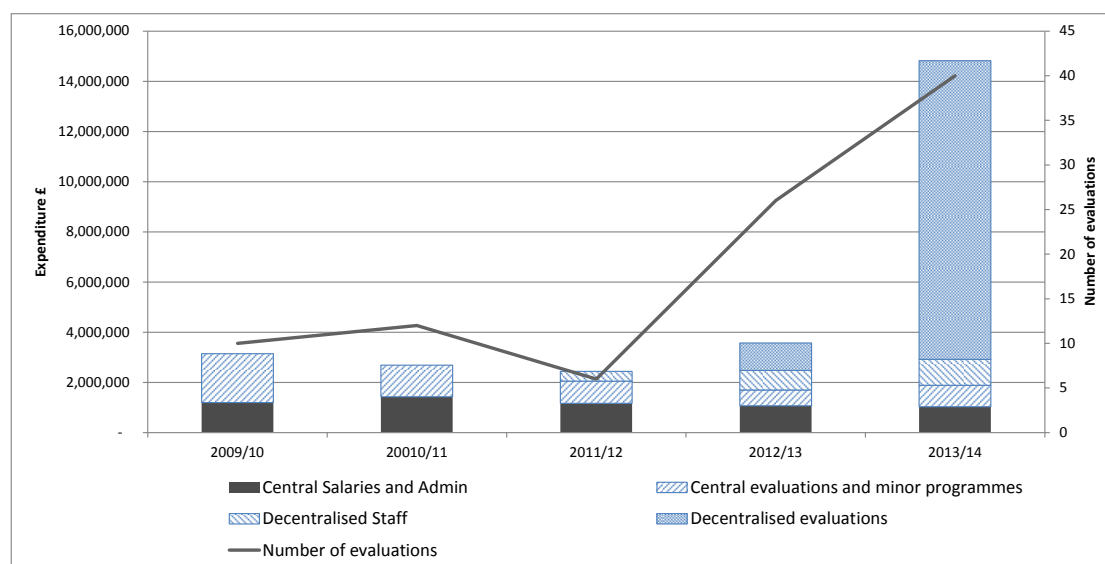
The average total cost to DFID of conducting an evaluation has fluctuated around £290,000[39] for the last five years. The average total cost per evaluation has changed little since 2009/10 as the steep increase in total evaluation related costs has been accompanied by a similar pattern for the number of evaluations published each year (see Figure 8.1). It could be argued that a fall in evaluation cost should have been realised due to economies of scale from conducting more evaluations. However, it is unlikely that such economies of scale could be realised during the early implementation of the embedding approach. They may yet be realised later in the process.

---

[36] E.g. synthesis reviews funded by the Strategic Evaluation Fund.
[37] Excluding strategic evaluation programme costs which are covered below.
[38] Source: Aries, EvD forecasts, review team estimates.
[39] When calculated in terms of total evaluation costs (central and decentralised), divided by the number of evaluations published each year.

There are limitations in the use of the total cost measure for evaluation cost, for example, due to time lags between appointment of staff and the publishing of evaluations. The figure is, however, consistent with the estimated direct cost of evaluations calculated in Section 8.3 below. It is anticipated that the average total cost per evaluation will increase after 2013/14 as the direct costs per evaluation are forecast to rise over the coming period (see Section 8.3 below).

Cost of strategic evaluation programmes

EvD has invested in a number of strategic evaluation programmes in order to improve the quality and relevance of IE within and outside DFID and to develop the capacity of evaluation partners, particularly in developing nations. Expenditure by EvD and other DFID departments in this area of 'Strategic Evaluation Programming' has significantly increased from £2.8 million in 2009/10 to £25 million in 2013/14. Strategic evaluation programmes relate to a range of programmes including CHAI and Global Partnership Development. They include a significant focus on advancing IE through investments in SIEF (IE to human development), 3ie and the proposed Centre of Excellence (CoE) in IE. Figure 8.2 provides an overview of the evolution of expenditure on strategic evaluation programmes by DFID departments.

Figure 8.2 Strategic Evaluation Programmes[40]
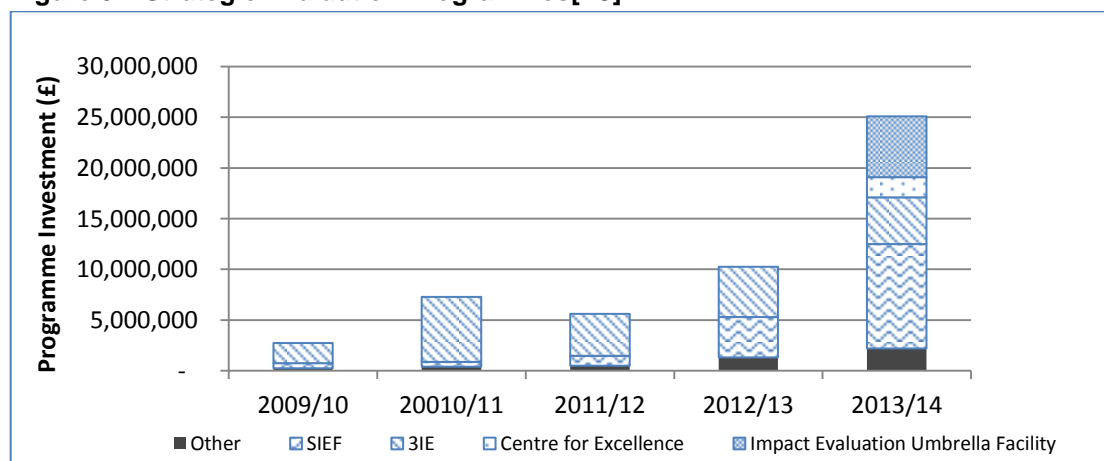
The increased investment in these strategic evaluation programmes reflects DFID's focus on driving the results and evaluation agenda with our partners and the wider development community. It also reflects the increase in programme spend across DFID during recent years.

The results and value for money of DFID's investments in these programmes are assessed through ARs. Individual evaluations of these programmes are not within the scope of this review of embedding evaluation.

---

[40] Source: Aries.

8.3 Economy and efficiency of independent evaluation

Direct evaluation contract costs

Direct evaluation contract costs are the amounts paid to external evaluation service providers and, unlike the calculation of total cost above, do not include DFID salaries and overheads. The median direct cost of evaluations has increased from £181,000 in the period prior to 2011 to a median cost of £300,000 from 2013 onwards. As can be seen from Figure 8.3, this is partly explained by an increase in the number of IEs[41] that will rise from two in 2012/13 to 20 in 2014/15. The median budgeted cost of IEs is higher than that of non-impact evaluations at £500,000. The median budgeted cost for non-impact evaluations from 2013 onwards is £200,000. This is comparable to the median cost of £181,000 during the period prior to 2012, during which few or no IEs were conducted.

Figure 8.3: Median Cost and Number of Evaluations[42]
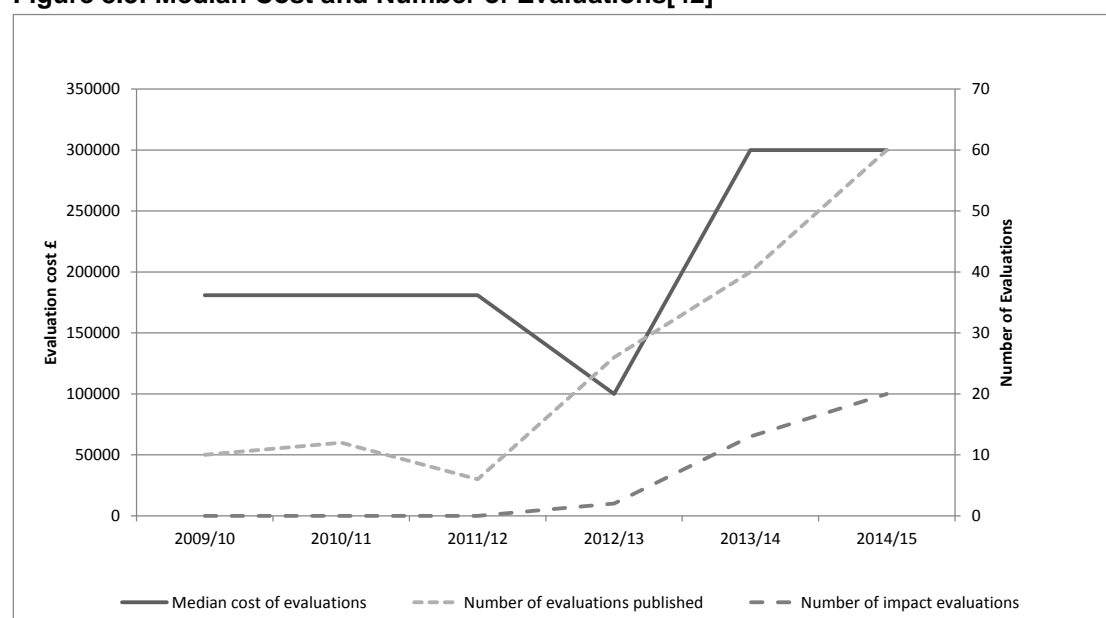
The cost of evaluations relative to other donors is difficult to assess due to the lack of available data and differences in programme portfolios and definitions of evaluation between donors. However, two approximate benchmarks are provided by the Danish (DANIDA) and Swedish (SIDA) development organisations: DANIDA estimates that it typically spends between £200,000 and £350,000 on the evaluations of its programmes and SIDA estimates that its larger evaluations can cost up to £500,000[43]. These indicators suggest that the forecast median direct cost of DFID non-impact evaluations at £200,000 is at the low end of the costs paid by these two donors. Meanwhile, the budgeted median cost for IEs is at the high end.

---

[41] Included in IEs are quasi-experimental, experimental and broader methods of IE.
[42] Source: EvD evaluation database, EvD estimates and actual expenses paid for 2012/13.
[43] Source: DANIDA and SIDA EAs.

The cost of IEs is in line with estimates from the World Bank's Development Impact Evaluation group, which estimates the average cost of an IE at about £500,000 ($800,000).

Table 8.1 – Evaluation Cost as a Proportion of Programme Value

	Median Cost (%)	Average Cost (%)
IEs	2.6	4.2
Other evaluations	1.5	3.3
All evaluations	1.9	3.6

Source: DFID EvD evaluation database

Evaluation costs as a proportion of programme value

DFID expects to spend a median of 1.9% of the programme value on independent evaluation based on budgets from evaluation commissioners. IEs are budgeted to cost 2.6% of programme value and other evaluations 1.5% of programme value. This is in line with the expectations of country evaluation strategies, which forecast expenditure of 1-4%. The mean cost of evaluations is higher due to the inclusion of a number of research related evaluations that represent a significant proportion of the programme value. This is due to the smaller pilot nature of the programmes and use of expensive evaluation approaches that rely on primary data collection.

The median cost of IEs in relation to programme value appears reasonable when compared with external benchmarks. The MCC estimates that it spends between 2-3% of project value on IE and the World Bank estimates the median cost to the bank at 1.4% (Independent Evaluation Group 2013). (This excludes other contributions made to World Bank IEs by project teams and partner governments.)

However, the cost of IEs varies across thematic sectors. A further detailed analysis of IEs in the education sector reveals an average forecast spend of 4.6% across 19 programmes that are expected to be evaluated.

Evaluation consultant daily rates

Consultant day rates provide another measure of the value for money of DFID's procurement of independent evaluations. An average daily evaluation consultant rate contracted by DFID has been calculated based on the procurement data from the GEFA. This daily rate is impacted by the seniority of staff required for DFID's evaluations and is not easily compared with day rate benchmarks of other organisations that procure different types of evaluation and consultant. Initial indications suggest that the average DFID rate is towards the higher end of the pay scales of other organisations such as the World Bank and Nordic donors. This may reflect a number of factors including the nature of the evaluation and the fact that it compares DFID's actual costs with other organisations' benchmark costs.

However, the higher than benchmark costs are in line with concerns regarding the shortage of supply of evaluators to meet the increased demand which DFID is placing on the market. They may also reflect the way in which evaluations are procured through the GEFA, which has seen limited competition for a large proportion of tenders (according to the recent GEFA Review carried out by Procurement Group).

9. Revisiting the ToC

9.1 Activities

All the activities outlined in the ToC have taken place to some extent or another.

EvD has been restructured and moved to RED and ICAI has been established. There are now 30 EAs providing support in decentralised posts. However, this figure is lower than the 40+ initially envisaged. What is more, the majority of those positions are hybrid posts with statistics and/or results, meaning that the overall level of decentralised evaluation support is considerably less than in the initial vision. A professional cadre has been established and now has 153 members.

Capacity has been built both with in-house training and, at higher levels, through agreements with 3ie and UKES. In addition, COs are carrying out their own training with the support of EvD. Evaluation is being financed by programmes and incorporated into management cycles. In many respects this has surpassed expectations, at least in terms of quantity of evaluations planned. But there have been comments from staff that in some cases, evaluation can be squeezed from the original budget proposed.

In terms of building global development evidence, DFID has committed significantly more funds to support international organisations such as 3ie and J-PAL to provide public goods. Synthesis studies are being developed internally which will also contribute to global development evidence. However, this is still at an early stage and needs more attention.

Outreach and communications are also areas where more can be done. Individual COs are working well with their partners, but DFID has lost some visibility in the eyes of some of its global partners. Internally a number of communication products have been developed, some specifically for evaluation, others using more general communication channels. However, these need to be reinforced and better publicised.

In the provision of routine support, EvD advisers have responded to requests, as do advisers in central divisions and the geographic central units. The SEQAS and GEFA have been set up to support decentralised evaluation through QA and through, in theory, making the contracting process easier. A routine reporting system has been established to ensure that EvD is aware of planned evaluations and their status.

An Evaluation Handbook has been produced and in 2013 the Evaluation Policy was updated to reflect the new structures.

9.2   Outputs

All the outputs have been achieved. These were never quantified, so it is difficult to assess how much the quantity and quality of the outputs has been affected by the relatively low amount of EA availability, compared to the original vision.

The aim of more staff in operational and policy divisions able to undertake specialist work on evaluations has been only partially achieved. A number of programme advisers and managers are now accredited to the cadre. However, there are indications that focus up until now has been on development and design. Whether there is the level of expertise necessary to manage a highly technical IE is debatable. The use of SEQAS has been beneficial, as has, in some circumstances, external steering groups.

Establishing a professional cadre of EAs has been achieved, though the cadre is broader than EAs. This has undoubted benefits but the question has been raised as to whether activities need to be focused on the higher levels of the cadre.

Renewed efforts are needed to develop the capacity of staff in DFID to manage evaluations. Participants at the PDC highlighted this as a key area for further development. As noted earlier, this is one area where there may need to be more focus in future.

Evaluations are commissioned and managed by spending units. Systems and platforms for sharing lessons and evaluation evidence have been developed, though mostly focusing on the internal audience. Although evaluations are published on the DFID website and Development Tracker, it is not easy for the global audience to find reports, as they are linked to projects rather than appearing automatically on an evaluation or thematic webpage.

Guidance notes have been developed. An Evaluation Policy was produced in 2013 (DFID 2013) and many COs and policy divisions have developed evaluation strategies. However, these are not required, though strongly recommended. The Africa region has taken a committed position on this, as is reflected by the much larger number of evaluation strategies in African COs.

9.3   Intermediate outcomes

The link between output and outcome is the most likely to be affected by uncontrollable factors or misplaced assumptions. Most of the intermediate outcomes have been achieved, at least formally, not least because evaluation is now an integral part of the BC. However, there are indications that in some parts of the organisation this is more of a checklist approach where staff do not always have time to consider fully the case for or need for evaluation.

Rigorous evaluation is a routine and accepted part of the programme cycle, though there is some question over how well the feedback mechanisms work (if they do

exist). Evaluation is possibly less routine for the policy cycle, though it may be too early to assess this.

The potential for increased use of robust and appropriate evaluation designs, approaches and methods is present. Evaluations are being planned from the start of a project rather than coming in as an afterthought, so there should be more baselines measured where necessary and there should be more awareness of different approaches to evaluations. However, the number of evaluations categorised as IEs in the EvD evaluation database indicates considerable ambition, which may be more than can be delivered.

An expansion of decentralised evaluations where the evidence base is particularly weak, and for larger programmes and more innovative programmes, is needed. Where there are evaluation strategies, many of these set out the factors given above as important elements to be considered when deciding what projects to evaluate. It is too early to assess properly whether this is actually happening. There are definite examples where innovative (or even controversial) programmes are being evaluated. However, it is difficult for a CO in isolation to identify those areas where evidence is weak, unless staff know what the broader evidence base is and what is being or has been evaluated in other offices and by other development partners. This is a weakness in some decentralised units.

When compared with the pre-embedding era, DFID staff and partners do have more capacity to manage and conduct high quality evaluations, and it is more widely spread across the organisation. In COs, staff work with partners to help develop their capacity, whether formally or through working together. As long as there is high level support to evaluation in DFID, this is likely to increase.

There are also clearer standards on evidence when new programmes are commissioned: The BC process promotes this. It should be expected that the increased number of evaluations will provide some of this evidence in the future, and analysis of current management responses to recently completed evaluations give some good examples of where this is happening.

9.4   Final outcomes

It is more difficult to assess the extent to which final outcomes have been achieved. These are difficult to measure and the following observations are based to a large extent on individual perceptions, of both the review team and the informants they interviewed.

Does DFID have a culture where staff routinely use evidence and evaluations in designing new programmes? This is undoubtedly the aspiration of most staff interviewed but there are barriers to achieving this which vary according to an individual's position. Staff often do not have time, not just for evaluations but for consideration of the implications of evaluative findings. EAs are not always brought in at the programme design stage. Senior staff may not give sufficient priority to the development of evaluative evidence, or understand the time it can take to produce

high quality evaluations. All of these can restrict both the production and the use of evidence and evaluations.

More evaluative evidence is becoming available, but not always in the areas where it is most needed. The review identified areas such as wealth creation and governance which are less well covered in terms of existing evidence but also in terms of planned evaluations. The relative autonomy of COs and the absence of central guidance, whether from EvD, RED or policy divisions, makes it unlikely that these gaps will be systematically filled. There is no clear incentive for COs to focus on the provision of what could, in part, be a public good. It is possible that there is a lack of awareness of the gap at country level.

Understanding of evaluation has increased considerably throughout the organisation, though some would say not by enough. There is a perception in some quarters that the demand for evaluation has waned somewhat recently. Dissemination of interesting and useful results would go some way to addressing this, which in turn makes the choice of what to evaluate more critical.

DFID subscribes to international quality standards. EvD takes QA very seriously and the systems set in place with SEQAS are quite rigorous. Over time, they should result in an improvement in quality. There are questions as to whether the number of evaluations is stretching the capacity of a relatively small market of contractors, and anecdotal evidence that it can be difficult to find contractors for specific evaluations, particularly where there are security or particular technical issues. DFID does not implement evaluations itself, but manages external contractors. However, the market is slowly expanding so this may become less of an issue in future, particularly if the skills of local contractors improve. There is no evidence either way that decentralisation has affected independence. For more sensitive evaluations, appropriate governance arrangements can and are put in place, with external steering or reference groups which should provide some protection against potential political pressure. Transparency may be more of an issue, not consciously, but because the increase in numbers of commissioning units has, at times, led to less visibility of individual evaluations.

As was remarked above, the assumptions made are crucial to the robustness of a ToC. Some of these have proved to be optimistic in the case of the embedding process. As mentioned above, staff have not been recruited to the extent originally envisaged and there are questions over the capacity to commission and undertake high quality evaluations, though this is increasing. It is difficult to show as yet that evaluations affect the views of key stakeholders, in part because relatively few of the planned evaluations have been finalised as yet. Dissemination of results may also be an issue.

Perhaps the most important factors are continuing political support, the available time of evaluation staff and spending units' prioritisation of evaluations within programmes. These are linked. To increase time available, there has to be support from senior levels to create space for evaluation and a budget has to be made available for evaluation. Senior level support is also, to some extent, dependent on

the external political environment. Although embedding evaluation has had notable achievements, it is still only three years since the process began. Evaluation may not be sufficiently embedded to withstand a change in the political context, at least at its present level.

10 Key Issues and Implications

10.1 Achieving an appropriate balance

Between learning and accountability

It is easy to suggest that the role of ICAI is to provide accountability to Parliament and the UK taxpayer and therefore that evaluation within DFID should focus on learning. ICAI's website does indeed focus on its independence and scrutiny role, rather than learning. The vision statement developed by EvD/DFID emphasises the role of evaluation as producing evidence, particularly on impact, but it also mentions value for money. ICAI reports directly to the International Development Select Committee on a regular basis, while evaluation in DFID is now an integral part of the business model.

The review team asked a number of informants what they saw as the primary purpose of evaluation in DFID. The vast majority mentioned learning in their answer and placed it above accountability. Only one respondent focused on accountability. Evaluation is seen as a way of finding out what works and why. ICAI is seen more as an auditor, focusing on accountability. However, when staff were asked about the principal purpose of individual evaluations, accountability was mentioned more often. This was particularly true for large projects or projects with multiple funders. Learning was seen as important in the evaluations of small pilot projects and where there was a significant evidence gap. If projects were undertaken in a context which was regarded as risky, then accountability was seen as important.

Although there were exceptions, the general view seemed to be that there were a number of mechanisms for accountability, including the work of ICAI, and the system of reporting on results. Evaluations almost always had an accountability element. This was more important for some projects and contexts than others, but a good evaluation could provide lessons and evidence where accountability mechanisms could not.

At the PDC, participants were asked about their views on whether the purpose of DFID evaluations appropriately balanced generating information for learning, including building the evidence base, and ensuring accountability for results and spending. Of the 53 who gave a firm response (rather than not answering or answering 'don't know'), 30 said that there was too much emphasis on accountability. 12 said it was about right and four thought there was too much emphasis on learning. So it would appear that one of the elements of the vision for embedding evaluation, that evaluation should be about evidence and lesson learning, has been taken on board by DFID staff in the Evaluation Cadre, even if they feel that there is still some way to go.

Between decentralised and centralised evaluation

The embedding process has resulted in a significant increase in the evaluation capacity and decision making power over what to evaluate at both CO and division levels. It certainly appears to have increased ownership in a number of COs, but is the resulting balance right in terms of roles and functions? Are some of the issues which have emerged from the analysis caused by insufficient central oversight, an inappropriate balance of resources and a lack of clarity in the roles and functions of the centralised EvD, as opposed to the decentralised divisions and COs?

There are a number of observations that can be made. Firstly, according to the results of the KIIs there are very mixed views on the relevance and effectiveness of EvD. Many respondents see EvD as highly relevant and are very appreciative of the support they have received. EvD provides important backup, for example with QA or where there is a tension between EAs and contractors. The guidelines and papers produced are appreciated, although some find it difficult to find time to digest them properly. There can be a lack of clarity as to the functions of EvD. This seems to be particularly true of EAs who have been recruited from outside of government as part of the embedding process. Some staff, who have been involved with EvD for a long time, feel that it has lost its way and that its mandate has become less clear.

Some of the perceived gaps in coverage could be addressed if there was stronger coordination and guidance from the centre. Staff felt that there were too many evaluations being planned and that they were insufficiently strategic. Could this be improved if EvD were given a coordinating role, and/or would it result in a loss of ownership? Would it be better to involve the central regional offices in the process?

There is also the question of synthesis studies. EvD has started to undertake these, but should it be their role or that of the relevant division? Additionally, should EvD go back to commissioning evaluations directly in areas such as aid effectiveness? If so, what are the cost implications? There was no view expressed that the situation should revert to what it was pre-2010. Staff appreciate EAs in countries, and when programme managers ask for advice it is more likely to come from regional or CO EAs rather than from the EvD Evaluation Support Group (ESG) team.

In terms of capacity, the clear message coming from staff is that although capacity has increased substantially over the last three years, it has not kept pace with the ambition in terms of number of evaluations. It is not a simple matter of insufficient decentralised capacity or not enough capacity at central level. It is more an issue of ensuring a better match between planned evaluations and capacity on the ground, in terms of both EAs and programme managers. Some COs have explicit targets for supporting their staff to achieve accreditation. One approach to improving the balance would be to encourage more COs to follow suit.

For demand and use of evaluations, the issue is not so much one of balance between centralised and decentralised evaluation. Rather, the relevant matters are resources available at the decentralised level to access and use in evaluations, time, and ease of access. There is also considerable variation in the enthusiasm for

evaluation shown by senior management, which in turn affects how their staff perceive evaluation in relation to other demands on their time. This is something that it is difficult for individual staff to address, particularly in a decentralised system, but is an area where central divisions, not just EvD, could take action to reinvigorate the profile of evaluation. One of the difficulties is that there has been an inevitable gap in the output of evaluations as the new approach becomes established. The heavy investment has still to show commensurate output. However, no individual CO has the resources to communicate its output outside of the country or region. This has to be the role of the centre, both EvD and the relevant departments.

Overall, the main issue in terms of balance between centralised and decentralised evaluation is a need for greater strategic guidance from the centre, with a possible adjustment and re-enforcement of roles. Greater resources will most likely have to come from a reinvigoration of decentralised units, which will have to be stimulated by EvD and RED more widely.

10.2   Fitness for purpose

COs and other spending departments/units

The extent to which evaluation is serving the purpose for which it was intended is a matter of both intent and application. The decision of whether and what to evaluate is being decided predominantly within programme teams and not collectively by senior management. This, along with the ability to pay for evaluation financed under BCs, has led to a rapid and uneven rise in evaluations commissioned since 2010. This reflects decision making driven primarily by accountability concerns.

The decision over whether or not to budget and plan for an evaluation within a new BC is often made in the default mode, namely it will be safer to have some evidence of how the initiative fared than not. Accountability is one of the drivers of the decision to evaluate in the EvD guidance, but alongside other considerations such as the evidence base (where thin, consider it), size of programme, level of innovation and level of risk. This is not a negative. The shift to embedding evaluation within DFID programmes has moved evaluation closer to the locus of decisions, providing potentially a much more rigorous source of information for organisational learning and impact than previously possible through centralised studies. However, the outstanding question is whether decentralised units will make best use of all of these studies.

This, in turn, leads to the question of whether offices and units have the capacity and determination to ensure evaluation relevance and quality. Evaluation staffing (FTE EAs) is uneven. The number of FTE advisers envisaged in the design phase of the embedding approach has not been achieved. Where EAs exist, they are almost always split posts with results or statistics positions and the evidence collected for this report suggests that the results role dominates. As the organisation focuses more on improving programme management following the 'End-to-End' Review of 2013, it is likely those in shared results positions will be increasingly challenged to manage their evaluation work. Incentives for individuals to work on programme cycle

management related tasks will always outweigh evaluation unless evaluation becomes more central to programme management.

The advisory role can work in COs and spending units where sufficient resources and training are provided to programme teams (advisers and programme staff) to ensure that evaluations are judiciously selected and resourced, and where the EA is enabled to invest time in the higher profile, complex and impact evaluations. This requires a management focus on evaluation within these units, which will be incentivised by a stronger corporate requirement from the centre, along with evidence of the utility of evaluations emanating from programmes. The latter requires substantial attention to be paid by EAs to the utility of the evaluations commissioned.

In addition, a stronger recognition of the relationship between evaluation and results is required by DFID. Greater recognition would better support the mutuality of evaluation and results at country level, particularly with the increased focus on programme management. This, in turn, might necessitate more central support on complex and impact evaluations, including jointly managed initiatives, to better ensure quality and utility at the country and corporate levels.

EvD

Three big changes occurred in April 2011:
- the relocation of EvD from reporting to the DG of the Finance and Corporate Performance Department to RED
- the change in role from designing, managing and following-up on independent evaluations, to rolling out and providing a service function to those in the COs and spending units responsible for evaluations
- the change in type of evaluations commissioned, from predominantly country and corporate evaluations to project and impact evaluations.

Of these moves, the relocation and the change in type of evaluations commissioned have provided a shift away from evaluation's role in servicing corporate performance and accountability requirements, and towards the generation of thematic evidence for learning. It is unclear if these changes were intentional. These are big changes alone but, with the additional shift of decentralising the responsibility, EvD's service and support function was immediately at odds with that of the rest of RED, where centrally commissioned research remains the norm[44]. The strongest interface between EvD and other functions of RED, in principle, is the dissemination and uptake of evidence generated by the products, though even this is not a mandated responsibility of EvD but of decentralised units[45].

---

[44] While recognising the establishment of the South Asia, and recently East Africa, Research Hubs, they still represent a small proportion of the overall research budget of RED.

[45] The nature of the service function that EvD carries out has also reduced its performance visibility, where results are achieved elsewhere. Previously its performance could be measured directly in terms of numbers, quality and uptake of evaluation it commissioned centrally, all concrete and tangible measures. Following the shift, the measures are now numbers of evaluations supported, staff trained and the like – all valuable but harder to defend.

RED's primary objective was outlined in the OP 2011-15. It is to make DFID more systematic in using evidence as a basis for how best to reduce global poverty through commissioning research, robust evaluations of DFID programmes, statistics and engagement with policy makers. The role of EvD fits clearly within the framework of supporting systematisation in generating and using evidence from evaluation, but not through commissioning evaluations.

Since its move to RED, EvD has served two primary functions. One is establishing and maintaining the policy framework and systems for supporting decentralised evaluation. The other is providing a one-on-one technical support service to those designing and managing complex, impact and strategic evaluations. The latter is managed by one team, the ESG team, comprised of 5.6 FTE EAs. A team database tracks specific support to 249 current evaluations, representing over half of all the ongoing evaluations in DFID and all known complex, strategic and impact evaluations. While this represents support of a ratio of 44 evaluations per FTE adviser, in practice most advisers are reportedly supporting in the range of eight-ten evaluations at any one time. The Capacity and Quality Group, with 1.5 FTE EAs and 3 evaluation managers, focuses on drafting and maintaining policy, guidance and training material. It also provides support on QA and procurement issues as well as leading on external capacity building with partners.

Both of these functions are appropriate in the context of a strongly decentralised evaluation function. To better support decentralised evaluation there will need to be a look at the responsiveness of EvD to demands for support, the quality of this support, and ultimately the quality of products (as a proxy). The Evidence Survey conducted in 2013 found that despite a positive view of evaluation, 45% of those interviewed felt that evaluation advice and support on offer was only somewhat sufficient or insufficient to their needs. This cannot be distilled between EvD and EAs in the decentralised offices and units, though a number of interviewees were requesting further support from EvD EAs. Nevertheless, given the ratio of advisers to evaluations, it appears that support is spread too thinly at present.

While there is no question that EvD played a critical and effective role in the establishment of a decentralised evaluation system across the organisation, it now needs to reconsider what is needed to make this system effective and efficient in delivery. This includes identifying and supporting priority areas, balancing requirements with incentives, and working effectively through various channels of the organisation to ensure sufficient resources are allocated for staffing as well as evaluations and follow-up.

10.3   Implications and options

Although it was not in the scope of the review to make formal recommendations, a number of suggestions arising from the findings are proposed for further consideration. These are set out below.

Evaluation coverage

The findings show that gaps have developed in terms of coverage, geographically, thematically and in areas such as aid modalities. This is an almost inevitable result of the current situation of decentralised decision making without strong coordination mechanisms. This is an issue which ought to be addressed, but without undermining decentralised ownership. One option would be an overall DFID evaluation strategy, though it is not yet clear how that could be made effective. Additionally, there should be much stronger involvement of senior management in decentralised units over the decisions to evaluate. This would be supported by greater information as to evaluation gaps and ongoing evaluations. It would be preferable to the present situation, which often leaves the decision in the hands of programme managers.

Support to operational units

DFID has made efforts to build the capacity of decentralised units to design, manage and use evaluations. However, the QA of evaluation ToR, inception reports and final evaluation reports has identified gaps in current practice. In particular, there is a need to focus efforts on ensuring evaluations are improved by:
- understanding why and when to commission evaluation
- contextualising evaluations and engaging stakeholders in an appropriate and timely manner
- selecting and implementing appropriate evaluation approaches while ensuring reliability of data and validity of analysis
- analysing and clearly presenting information in a useful and timely manner.

These issues point to a critical need for DFID to ensure that operational units have the capacity to manage evaluations well, including engaging stakeholders during evaluations and undertaking QA of evaluative activities.

Variation in quality across regions, as indicated in QA reports, appears to reflect the distribution of EAs. This indicates a need to ensure countries and regions have adequate oversight and technical support to assure quality at each stage of the evaluation process. There is a similar unevenness of coverage across divisions and directorates. Where possible, EvD acts as a backstop for central and decentralised evaluations, but given the small size of EvD this has resulted in mismatched coverage of technical input and QA. This will only be addressed if two conditions are met. First, evaluations should be considered as part of good programme management, subject to the same due diligence and review, with good products lauded and used and poor ones flagged and addressed. Second, senior management in decentralised units should be convinced of the importance of evaluation and allocate resources accordingly.

Making better use of evaluation

There is a perception that enthusiasm for evaluation has waned since the embedding initiative was started in 2010. This is, in part, due to staff changes, particularly at senior level, but also because of a reduced visibility of evaluation results. EvD

stopped most of its centrally commissioned evaluations in 2010 and it has taken time for decentralised units to develop evaluation plans. Where evaluations have been completed, the focus has been on sharing results with key stakeholders, either at country level or in specific thematic areas.

It will be important for the DFID corporate centre, whether EvD or policy divisions, to communicate evaluative findings more broadly and at higher levels. This will provide assurance that the funding that is currently being spent on evaluation is making a difference to policy and programmes. Interest can only be maintained and increased if the use of evaluation is demonstrated.

Strengthening evaluation management

There has been considerable enthusiasm shown by programme managers for conducting IEs, which comprise 28% of planned evaluations. EAs have expressed some concern that this reflects a lack of understanding of what is required to conduct a good IE, in terms of time, resources and the technical capacity to manage such an evaluation properly.

Non-experimental approaches to evaluation, such as those shown in process and performance evaluations, have an important role to play in terms of providing continuous learning and improving programme design. However, to justify the significant amount of funding that could go towards these, there must be an improved system for drawing lessons from these evaluations and applying them. In some cases the importance of findings from these evaluations will be greater at local or regional level, and programme managers should be encouraged to focus on large, complex or innovative projects. These requirements are written into many of the evaluation strategies, but the extent of use of evaluation findings it will not be evident until these strategies are themselves reviewed.

Given the resources and technical capacity necessary to undertake IEs, an implication is that these should be fewer in number than currently planned. They can be managed from a decentralised level, and indeed there are useful examples already of IEs managed by COs. But these have needed quite extensive technical support, whether from EvD and RED, or from an external adviser or steering group. It will be important that COs develop a more intensive engagement with EvD and RED when they undertake these types of evaluation and that EvD and RED have the resources to provide more intensive support.

For this to happen, EvD will need to make certain decisions. Given the number and complexity of ongoing evaluations, there are several options:
- staffing in EvD is increased
- the responsibility is shared with other parts of RED
- the numbers of evaluations across the organisation are dramatically curtailed
- a much greater investment in staffing is made at the decentralised level.

Enhancing evaluation governance

There is a need to strengthen the governance arrangements that determine DFID's strategic approach to evaluation. The current EESG has fulfilled its initial purpose and now requires a new mandate, more senior membership and a clearer reporting line to the EMC and/or the IC. The mandate should reflect the two bodies that evaluation serves – the body corporate, on accountability for results and organisational lessons, and the body of evidence, through generating new knowledge on what works in development practice. To do so, the mandate of the EESG would include commissioning the development of a DFID evaluation strategy. This would specify priority sectors, geographies and thematic evaluation areas, as well as clear guidance on decisions on what to evaluate. The new mandate would also give the evaluation sub-committee responsibility for advising DGs and directors of any changes required in their divisions for effective implementation of the strategy. The membership would include directors of geographic divisions and policy divisions to ensure organisational commitment and action in response to requests from EESG.

The strengthening of overall evaluation strategy will address two key findings in this report. Firstly, the review has identified variable regional and thematic coverage and the minimal involvement of relevant policy or evaluation teams in decisions to evaluate. Secondly, there is a need to reinvigorate decentralised engagement with evaluation. A more strategic approach to evaluation has the potential to ensure optimal use of evaluation resources. It can thereby fulfil the overall purpose and function of evaluation as an essential tool for increasing the effectiveness and impact of development assistance.

Annex 1. Review Team and Governance

The governance and implementation of REED was designed to balance the competing requirements of speed, independence and value for money. The review was managed by an internal Review Working Group that was supervised by the Evaluation Department Senior Management Team (EvD-SMT) and a Reference and Advisory Committee (RAC).

The roles, responsibilities of those involved in the review were as follows:

- **The Review Working Group** planned and implemented the review under the strategic guidance of the EvD-SMT and following the technical advice of the RAC. The working group was comprised of one external expert in the lead role and five members of EvD (part-time). An extended group of individuals also contributed to the review, as listed in the acknowledgements.

- **The EvD-SMT** was responsible for overseeing the Review Working Group. The EvD-SMT had ultimate responsibility for the review.

- **The RAC** provided:
  o   expert inputs on the design and draft outputs of the review
  o   advise on quality of the products
  o   suggestions on recommendations and uptake of the findings within and beyond DFID.

The RAC comprised individuals external to the organisation, but who are familiar with centralised and decentralised evaluation, and internal individuals who are decision makers with an overview of the embedded evaluation system. The committee was composed of:

- Penny Hawkins (Chair), Head of Evaluation, DFID
- Susan Ulbaek, Head of Evaluation, DANIDA
- Sandra Baldwin, Deputy Head, DFID Southern Africa
- Martin Dawson, Deputy Head, DFID Malawi
- Hans Lundgren, Head of DAC EvalNet Secretariat
- Julia Kemp, Head, DFID East Africa Research Hub
- Ellen Wratten, Deputy Head, Policy Division, DFID.

Annex 2. Review Approach

Review	Sub-sections	Data Sources	Analysis and Outputs
1. Embedding strategy and vision	• **Timing and context of embedding** • **Embedding vision and implicit strategy** • **Embedding ToC**	- Document review - KIIs - ToC workshop	- Description of context of embedding decision - Objectives and assumptions of embedding - Description of strategic plan - ToC
2. The embedding process	• **Steps taken to embed** • **Governance** • **Capacity building** • **Professionalisation** • **QA**	- Document review - KIIs	- Overall timeline - Analysis of staffing 2010-13 - Capacity building activities 2010-13
3. Preliminary assessment of embedding effectiveness	• **Quantity of evaluations**	- ARs - EvD programme manager's database - 2011 ICAI review of DFID evaluations	- Analysis of number and coverage of evaluations pre-2011 and post-2011 - Draft chapter section
	• **Quality of evaluations**	- Country evaluation strategies - SEQAS reports - KIIs -Case studies	- Analysis of understanding of quality - Analysis of practice of quality
	• **Coverage**	- Country evaluation strategies - ARs - EvD's database - 2011 ICAI Review of DFID evaluations - KIIs	- Analysis of geographic coverage - Analysis of thematic coverage - Types of evaluation
	• **Demand and use**	- Country evaluation strategies - Management response to strategies - Evidence Survey - KIIs - Focus Group Discussions (FGDs) at PDC	- Analysis of country evaluation strategies
	• **Capacity**	- Staffing reports - Case studies of offices - FGDs at PDC	- Analysis of existing capacity - Analysis of how capacity is being built
	• **External perception**	- KIIs	
4. Lessons learned since 2010	• **Balance between learning and accountability, centralised and decentralised**	- KIIs	
	• **Value for money**	- Financial data linked to output data over period 2007-13	- Trends in financial data - Costs of evaluations - Cost of support function - Cost of QA
	• **Strengths and Weaknesses**	- Based on the above analysis	

KIIs were an important part of the methodology. Internal interviewees were purposively selected across the range of locations, sectors and functions. In order to understand better the genesis of the embedding evaluation process, the review team also interviewed the 'architects' of embedding. These were a number of individuals, both DFID staff and some external advisers, who were in positions of influence in the period 2009-10 and played an important role in how the embedding process unfolded. Some external informants were also interviewed to give an outsider perspective on the changes.

The table below gives details of the numbers of people interviewed.

List of Interviewees	
Category	**Number**
Architects of embedding evaluation	**9**
Internal DFID staff members:	32
• Country based staff	14
• Heads/deputy heads of COs	5
• Evaluation/results advisers, country based	4
• EvD staff	3
• UK based senior staff[46]	9
• UK based evaluation/results advisers (not EvD)	4
• Members of the Evaluation Cadre	16
• Members of the Evaluation Cadre, not EAs	4
External to DFID	3
Total number interviewed	**44**

Protocols were developed for the three different categories of interviewee (architects, internal and external) and the responses were entered into a spreadsheet by question to enable a systematic analysis of responses. Because of time constraints, not all questions were answered by all respondents.

---

[46] HoP, deputy head, deputy director and director.

Annex 3. Documents Reviewed

Internal documents used for content analysis:
DFID (2009). Building the Evidence to Reduce Poverty: The UK's policy on evaluation for international development
DFID: EESG. Minutes of meetings for period covering: 4/6/10 until 26/2/13DFID: EvD Back to Office Reports covering 20 country visits between March 2010 and March 2013
DFID: Professional Development Conference proceedings for: September 2011, September 2012, and November 2013
DFID (2010). Vision for DFID Embedding Evaluation. Approved by EESG
DFID (2011). EvD Operational Plan 2011-2015. EvD
DFID (2011). Evaluation Stock-take. EvD
DFID (2012). Paper to the Investment Committee on DFID Evaluation. EvD
DFID (2012). How to Note: Global Evaluation Framework Agreement (GEFA). EvD
DFID (2012). How to Note: Specialist Evaluation & Quality Assurance Service. EvD
DFID (2013). 'Fit for the Future: Joint Evaluation and Statistics Cadres Professional Development Conference'. 11th-15th November 2013, London. DFID
DFID (2013). Country Office and Departmental Strategies. DFID Insight: http://epe-insight/evaluation/policy%20and%20strategy/Pages/Country-office-and-departmental-strategies.aspx
SEQAS (2013). SEQAS Annual Review 2013: Online Survey. DFID
Works cited:
Cabinet Office (2010). The Coalition: Our programme for government. May 2010. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/78977/coalition_programme_for_government.pdf
DFID (2010a). Embedding evaluation in DFID - Initial scoping note. July 2010.
DFID (2010b). Vision for DFID Embedding Evaluation. Approved by EESG.
DFID (2010c). Minute to Directors on ICAI and Embedding Evaluation. September 2010.
DFID (2010d). Question and Answer briefing on evaluation.
DFID (2012). Final Review of the Quality Assurance Process for DFID evaluations: End of Contract December 2012. http://dfidinsight/stellent/websites/insight_quest.asp?txtDocID=3841717
DFID (2013a). DFID Evaluation Policy 2013. 30 May 2013. https://www.gov.uk/Government/uploads/system/uploads/attachment_data/file/204119/DFID-Evaluation-Policy-2013.pdf
DFID (2013b). DFID Evidence Survey Results Report. 30 November 2013. http://r4d.dfid.gov.uk/Output/194821/Default.aspx
DFID/ICAI (2011). Memorandum of Understanding between the Department for International Development and the Independent Commission for Aid Impact. http://icai.independent.gov.uk/wp-content/uploads/2011/05/May-MOU-DFID-ICAI-_final_1.pdf
Drew, R. (2011). Synthesis Study of DFID's Strategic Evaluations 2005 – 2010,

produced for the Independent Commission on Aid Impact.
http://icai.independent.gov.uk/wp-content/uploads/2010/11/Synthesis-Study-of-DFIDs-Strategic-Evaluations-2005-201011.pdf

Grant, E. and M. Chiche (2013). DRAFT 'Review of the Governance and Security Pillar Evaluation Strategy: Where are we one year on from the launch?' DFID

IACDI (2007). Minutes of IACDI meeting. 6 December 2007.
http://collections.europarchive.org/tna/20090908104656/http://iacdi.independent.gov.uk/wp-content/uploads/iacdi-minutes-dec07.pdf

ICAI (2013). Triennial Review of the Independent Commission for Aid Impact Completed. 17 December 2013. http://icai.independent.gov.uk/2013/12/17/triennial-review-independent-commission-aid-impact-completed/

Independent Evaluation Group (2013). World Bank Group Impact Evaluations: Relevance and Effectiveness. February 2013.
http://elibrary.worldbank.org/doi/book/10.1596/978-0-8213-9717-6?queryID=183%2F1915835&

Riddell, R. (2009). The Quality of DFID's Evaluation Reports and Assurance Systems. November 2009.
http://webarchive.nationalarchives.gov.uk/20101216162913/http://iacdi.independent.gov.uk/2009/12/evaluation-quality-Review-synthesis/

York, N. (2011). 'Management Board Paper'. 20 April 2011. DFID

SEQAS reports:

Entry ToR:

1024 Tanzania QA of ToR – Education Quality Improvement Programme, Tanzania. (EQUIPT) IE

1026 Ghana ToR – Evaluability Assessment for a Ghana Adolescent Reproductive Health Programme

1036 ToR for evaluation manager – Global Trade and Global Value Chains Initiative

1037 Sierra Leone QA of ToR – Improving Schools in Sierra Leone

1061 Africa QA ToR – Africa Enterprise Challenge Fund

1064 India QA – Draft ToR development and implementation of a strategy for independent IE of the Odisha Girls Incentive Programme

1072 Somalia ToR – DFID Somalia Humanitarian Programme 2013-17

1088 Global QA of ToR – Water, Sanitation & Hygiene (WASH) Results Programme
1118 South Sudan QA of ToR – Health Pooled Fund

Entry Non-ToR (inception):

1093 Zambia QA/review – Evaluation Design: Community Health worker evaluation Zambia

1107 Africa Evidence and Action for Maternal and Newborn Health Policy in Sub-Saharan Africa

1109 QA of inception report – India Impact Evaluation of 'MAMATA'-Odisha Maternity Support Scheme

1031 Zambia QA of inception report – Zambia Sanitation Evaluation

1103 Burma Independent Evaluation of Artemisinin Monotherapy Replacement in the Private Sector to Support the Containment of Artemisinin Resistant Malaria in Burma

1110 South Asia QA of inception report – Mahila Samakhya- Education for Women's Equality Programme

1121 Bangladesh QA of inception report for complex evaluations – Accelerating Improved Nutrition for the Extreme Poor In Bangladesh

Exit:

1014 Sudan Evaluation of DFID Sudan Humanitarian Bilateral Projects

1019 Africa Investment Climate Facility for Africa – Review Report

1028 Tajikistan Central Asia Regional Migration Programme

1040 Global Trade Facilitation Facility

1052 Ethiopia QA WASH Capacity Building

1057 Asia South Asia Food & Nutrition Security Initiative

1081 Zambia Child Grant Programme

1105 Ghana Election Security Programme

1117 Afghanistan Humanitarian Assistance through Mobile Cash Transfer in Northern Afghanistan

1122 Vietnam HIV/AIDS Programme

Annex 4. Timeline of Embedding Evaluation

Date range	Key event	Comment
2007	IACDI formed	Established to guarantee the independence of the evaluation function in DFID and the use of evaluation results to enhance the delivery and impact of UK development assistance.
June 2009	Publication of the first DFID Evaluation Policy: 'Building the Evidence to Reduce Poverty – The UK's policy on evaluation for international development' (June 2009)	Reflected that "DFID will seek to expand and sustain a strong culture of decentralised evaluation across the department". The accompanying ten-point policy implementation plan noted that DFID will "significantly increase the number of decentralised evaluations in programmes and projects".
September 2009	IACDI report released: 'The Quality of DFID's Evaluations and Evaluation Management Systems'	Report produced for IACDI by Richard Manning and Burt Perrin (Roger Riddell was the IACDI lead for this report). The quality review study looked at both central and decentralised evaluations in DFID and made comparisons to similar work in five peer bilateral institutions.
Effective from 1 April 2010	Treasury: agreement that all externally contracted evaluations can be funded from programme budgets	Prior to 1 April 2010, evaluations of partner programmes could be funded from programme budgets, **but** evaluation of DFID policies and programmes counted as an administrative cost. From 1 April 2010, all externally contracted evaluations, whether of partner programmes or DFID, could be funded from programme budgets.
May 2010	Coalition government formed	Promises to deliver greater transparency and scrutiny of aid spending to deliver value for money for British taxpayers and to maximise the impact of the aid budget.
June 2010 onwards	Two initiatives underway: i) establishing the independent aid watchdog (ICAI) and embedding evaluation within DFID ii) EESG established	Initiatives were to establish an independent aid watchdog (ICAI) (led at DG level) and to embed evaluation within DFID (led at DG level). Initially there were two separate steering groups for the two work-streams but these were merged. The combined EESG then met fortnightly but with ICAI issues addressed monthly. EESG established to manage the process of developing a culture of evaluation within DFID meetings – held every two weeks initially, changed to monthly from October 2010.
2010 onwards	In-house Principles of Evaluation training available	Series of in-house training courses held in HQ and COs.
2010	Series of key reviews underway across DFID	Multilateral Aid Review, Bilateral Aid Review, Humanitarian Emergency Response Review.
2010 to March 2013	Around 20 country visits completed	Series of outreach engagement visits to COs to promote the embedding evaluation initiative and provide training.
23 June 2010	EESG agrees to launch stock-take of decentralised evaluations	Planned stock-take of decentralised evaluations (put on hold during the pre-election period) agreed to proceed, to cover the whole of DFID.

July 2010	Vision for DFID on embedding evaluation agreed	• World class in using evidence to drive value for money… • Programme design driven by rigorous evaluation of what works…. • Take measured risks using high quality evidence…. • Help partners to generate/use evidence, build capability
July 2010	Joint paper (EvD/RED) on putting evidence into use	Tabled to the Development Policy Committee.
29 July 2010	MB discussion on evaluation	Main points of paper: - scope and ambition- location of residual central evaluation work within DFID- discussion of whether any joint evaluation work should remain in-house.
September 2010	First board meeting on evaluation accreditation held	The decision to develop a professional cadre for evaluation was taken before decentralisation so the initial 'vision' was built around the existing centralised function.
10 September 2010	Note to directors from DGs	Minute recaps the overarching objective of the embedding evaluation initiative and outlines the implications for divisions
September 2010 onwards	Series of meetings with directors	Following the note to directors, a series of meetings is held with directors to discuss EE policy and placement of new EAs (additional 40 FTE EAs planned).
October 2010	Presentation to HoPs/Cabinets on EE and planning	
7 October 2010	EESG discusses embedded posts	Recognises that a lot of work is to be done around identifying posts and recruiting. Phased approach required for assessments, etc.
October 2010	ICAI Chief Commissioner announced	Graham Ward announced as ICAI Chief Commissioner.
October 2010	IACDI's last formal meeting held	
November 2010	EvD restructure process underway	EvD staff apply/compete for posts under the new structure.
2010/11	Evaluation strategies under preparation	Range of evaluation strategies being prepared, some of which are tabled to EESG for advice and information.
14 January 2011	ICAI launches 12-week public consultation	Opportunity for the public to comment on which areas of UK overseas aid they would like ICAI to cover.
April 2011	EvD joins RED	EvD transfers to RED, specifically the EE pillar. The evaluation chief professional officer/head of evaluation reports directly to the director of RED.

April 2011	EvD's OP 2011-15 produced	
April 2011	Head of EvD's paper to MB on evaluation	Issue: strengthening and formalising the DFID process for setting the strategic direction/DFID's mechanism for follow up on evaluations/oversee DFID's Evaluation Policy update.
2 May 2011	EvD moves to its new operating model	Organogram of 8 November 2010 identifies 16 staff (reduction from 22).
12 May 2011	ICAI official launch	ICAI officially launched and work-plan released.
June 2011	EvD hands over responsibility to ICAI	EvD transfers responsibility for commissioning new external evaluations and reviews to ICAI and to DFID operational units for decentralised evaluations (those commissioned by staff responsible for managing DFID's programmes, policies and partnerships or by their development partners.
June & October 2011	External recruitment at assessment centres held	EESG kept informed of the recruitment progress.
September 2011	First Evaluation PDC (London)	Evaluation network holds its first event in London. Key aim is for the cadre to network and identify main issues for working together and what will help to support the function going forwards. Includes one day UKES/DFID workshop: 'Development & Evaluation: Practical Ways Forward'.
October/ November 2011	Evaluation stock-take completed	Baseline position for evaluations agreed.
By December 2011 (and onwards)	Four accreditation rounds completed	As of 13 December, 100 staff accredited at different levels (4 levels exist). Further accreditation Boards take place during 2012 and 2013.
January 2012	New AR and Project Completion Report templates introduced	Changes to increase emphasis on evidence, results and value for money. Important for staff to understand the distinction between these reviews and evaluation.
13 March 2012	EESG agrees that the Directors' Statement of Assurance (DSA) is to stipulate that evaluations should have management responses	DSA requires further strengthening.
April 2012 onwards	First in depth, technical training course delivered by the UKES	Course open to DFID staff, other government departments and UKES members.
June 2012	Paper to the IC on DFID's evaluation programme	IC encouraged by high level of evaluation activity but notes the challenges in moving to a more strategic approach, posed by DFID's devolved working practices, including managing risks with poorly conceived and implemented evaluations. Agrees that EvD will maintain a regular flow of information to the IC and produce an evaluation Annual Report in 2013.

26 June 2012	QA entry and exit evaluation products mandated	EESG accepts the EvD recommendation that QA should be mandatory.
August 2012	Michael Anderson, DG, no longer chair of EESG	
27 August 2012	GEFA established	Pre-qualified suppliers available to complete evaluations on behalf of DFID (and DECC/Defra for the ICF) (two-year framework).
September 2012	2nd Evaluation PDC (East Kilbride)	PDC theme: 'Achieving the Vision' (vision for evaluation and statistics, joint event with the Statistics Cadre).
9 October 2012	EESG meeting	EESG to continue as an 'evaluation leadership group', to be convened and led by the Head of EvD with the DG to attend on an ad hoc basis.
1 December 2012	SEQAS established	Consortium selected to provide specialist evaluation advice and support PLUS independent QA of all entry and exit products (three-year contract).
31 May 2013	New Evaluation Policy published	Original steer from the EESG/IC/Development Policy Committee was to 'refresh' the 2009 policy to reflect the establishment of ICAI and DFID's embedding evaluation agenda. In mid-2012, that view changed and a complete update was commissioned which involved wide internal/external consultation.
11 June 2013	EESG meeting	Papers on expanding the Evaluation Cadre (to include research). Update on the recent paper to the IC on planned evaluations and fit with DFID strategic priorities.
July 2013	2013 Annual Evaluation Report published	
September 2013	Procurement for CoE for IE advertised	OJEU advert placed for CoE. Tenders due by the end of November.
October 2013	REED launched	
November 2013	3rd Evaluation PDC held (London)	Theme: 'Fit for the Future' (joint conference with the Statistics Cadre).
13 December 2013	Draft REED report submitted to the head of evaluation	Draft report circulated for comment to the Evaluation Cadre and the RAC.
27 Jan 2014	Final Report submitted	

Annex 5. DAC Evaluation Quality Standards

Complete List of DAC Quality Standards for Development Evaluation:

1. Overarching considerations
1.1.   Development evaluation
1.2.   Free and open evaluation process
1.3.   Evaluation ethics
1.4.   Partnership approach
1.5.   Co-ordination and alignment
1.6.   Capacity development
1.7.   Quality control

2. Purpose, planning and design
2.1.   Rationale and purpose of the evaluation
2.2.   Specific objectives of the evaluation
2.3.   Evaluation object and scope
2.4.   Evaluability
2.5.   Stakeholder involvement
2.6.   Systematic consideration of joint evaluation
2.7.   Evaluation questions
2.8.   Selection and application of evaluation criteria
2.9.   Selection of approach and methodology
2.10.   Resources
2.11.   Governance and management structures
2.12.   Document defining purpose and expectations

3. Implementation and reporting
3.1.   Evaluation team
3.2.   Independence of evaluators vis-à-vis stakeholders
3.3.   Consultation and protection of stakeholders
3.4.   Implementation of evaluation within allotted time and budget
3.5.   Evaluation report
3.6.   Clarity and representativeness of summary
3.7.   Context of the development intervention
3.8.   Intervention logic
3.9.   Validity and reliability of information sources
3.10.   Explanation of the methodology used
3.11.   Clarity of analysis
3.12.   Evaluation questions answered
3.13.   Acknowledgement of changes and limitations of the evaluation
3.14.   Acknowledgement of disagreements within the evaluation team
3.15.   Incorporation of stakeholders' comments

4. Follow-up, use and learning
4.1.   Timeliness, relevance and use of the evaluation
4.2.   Systematic response to and follow-up on recommendations
4.3.   Dissemination