IACDI The Independent Advisory Committee on Development Impact

The Quality of DFID's Evaluations and Evaluation Management Systems:

How do they compare with other agencies?

Richard Manning
Independent Consultant

September 2009

Table of Contents

Table of Contents iii

Acknowledgements v

Executive Summary vii

1. Introduction 1

2. A Comparison of DFID Evaluation Studies with Studies by Comparator Agencies 3

2.1. Introduction and selection 3
2.2. Country Programme Evaluations 4
2.3. Thematic Evaluations 11

3. A Comparative Assessment of Evaluations Department's Management Systems and Processes for ensuring the quality of evaluations undertaken and commissioned by DFID, taking into account the practices of comparator agencies 21

3.1. Methodology and approach 21
3.2. Setting a favourable environment for high quality evaluations: how does DFID compare? 22
3.3. Deciding topics for evaluations likely to encourage relevance and buy in: how does DFID compare? 24
3.4. Involving partners and other donors in evaluations: how does DFID compare? 26
3.5. Managing evaluations with a view to a quality product: how does DFID compare? 28
3.6. Disseminating findings in a way likely to secure impact: how does DFID compare? 30
3.7. Following up recommendations systematically: how does DFID compare? 32
3.8. Outside Stakeholders' perception of evaluation products: how does DFID compare? 33
3.9 Handling evaluative material commissioned by other parts of the agency: how does DFID compare? 34

List of Annexes

Annex 1	Assessments of Comparator Evaluations considered in the Review	37
Annex 2	Aggregate Ratings - Non-DFID Evaluations	41
Annex 3	Norway - Zambia Evaluation Executive Summary	42
Annex 4	DAC Working Tool for Peer Reviews and Assessments: Evaluation Systems and Use	48
Annex 5	European Commission	60
Annex 6	Irish Aid	65
Annex 7	NORAD Evaluation Department	70
Annex 8	SADEV	75
Annex 9	SIDA Evaluation Department (UTV)	79
Annex 10	IOB (Netherlands)	83
Annex 11	Comparison of Resources for Evaluation: DFID and Comparator Agencies	88
Annex 12	Evaluation of the Commission's Support to United Republic of TanzaniaL Excerpt from 'Fiche Contradictoire'	89

Acknowledgements

I should like to acknowledge in particular the helpful guidance given me by Roger Riddell in this assignment and the co-operation and assistance provided by my fellow-consultant, Burt Perrin. Both have helped me considerably, not least from their much greater knowledge of evaluation.

I should also like to express my thanks to Nick York, Julia Compton, Helen Wedgwood and John Murray of DFID's Evaluation Department, who have provided useful advice and practical assistance. I also learned a good deal from DFID management and from staff of other interested DFID departments, notably Minouche Shafik, Mark Lowcock, Andrew Steer, Sam Sharpe, Alan Winters, Jim Drummond, Marcus Manuel, Jo Bourne, Liz Ditchburn, Sarah Cooke and Mike Smithson.

A particular word of thanks goes to the heads of the evaluation units of comparator agencies: Jean-Louis Chomel (European Commission), Fintan Farrelly (Irish Aid), Bram van Ojik (IOB/Netherlands), Asbjorn Eidhammer (NORAD), Gunilla Tornqvist (SADEV) and Joakim Molander (Sida), all of whom gave generously of their time and experience. I should also record my thanks to Bram for arranging a set of meetings also with other interested officials in the Netherlands Foreign Ministry. Ole Winkler Andersen kindly checked my references to DANIDA.

Finally, my thanks to Hans Lundgren and his Peer review Division colleagues at OECD and to Niels Dabelstein at the Danish Institute for International Studies for their interest and advice.

Responsibility for the report is however mine alone.

Executive Summary

Background

S1. This report addresses two issues. Firstly, how the quality of DFID's evaluations compares with those of a set of other donors. Second, how good are Evaluation Department's management systems for ensuring the quality of evaluations undertaken and commissioned by DFID, also looking at the practice of comparator agencies.

S2. In agreement with Roger Riddell, it was agreed that the comparator agencies would be Irish Aid (Evaluation and Audit Unit), Netherlands (IOB), Norway (NORAD Evaluation Department), Sweden (SADEV and Sida's Evaluation Department, UTV), and the European Commission (Evaluation Department).

Conclusions and Recommendations

S3. I have grouped my conclusions and recommendations into, first, the quality of the studies, and second, the quality of management.

Quality of the Studies

S4. It is impossible to make a soundly-based statement as to the relative quality of DFID evaluations in comparison to those of other agencies without a much more wide-ranging study than the present. For what it is worth, on the basis of my survey of some dozen evaluations carried out by comparator agencies, using the same methodology as that adopted by Burt Perrin for his linked study of the quality of DFID evaluations, I found a similar range from good to borderline on his quality scale. Based on this extremely limited sample, I saw nothing to suggest that there were systematic differences between the quality of evaluations commissioned by DFID and those commissioned by other agencies. This may reflect the fact that the agencies concerned use broadly similar practices and standards, and employ many of the same universe of consultants. There are nevertheless some differences which may be significant, and which are commented on below under Quality of Management.

S5. I looked specifically at the two main types of evaluation carried out by DFID's Evaluation Department, country programme evaluations and thematic evaluations.

A.1 Country Programme Evaluations

1. *Conclusion:* While all donor agencies need for accountability purposes and for future planning to have some way of assessing at least the effectiveness of their country strategies (if not the impact of their country programmes), the way in which they do so is quite varied, ranging from internal reviews to light evaluations as practised by DFID, more strongly resourced ones such as those of the EC (or the World Bank), or the occasional 'long-period' study as tried at different times by Sida, NORAD and DANIDA. Among the agencies included in this study, half of them (DFID, the EC and Irish Aid) all carry out country programme evaluations on a regular basis, and half (Netherlands, Norway, Sweden) do not.

It is not clear that one approach is inherently superior to another, though both DFID and EC evaluators would argue that their models provide consistent and independent approaches which give relatively high coverage of the bilateral programme. Donors need to consider how their accountability and assessment needs can be met in a cost-effective but sufficiently robust manner. More complex and expensive evaluations may not offset the cost and time incurred. The quality of data and the adequacy of frameworks for monitoring and evaluating country programmes are significant issues for comparator agencies as well as for DFID.

Recommendation: DFID management, in collaboration with Evaluation Department, should consider replacing its present 'one-size-fits-all' system for country programme evaluations (CPEs) with some deliberate experiments, including one or two better resourced CPEs, some simpler reviews and (see 3 below) more host-country-led approaches.

2. *Conclusion:* Many of the lessons (for example about maintaining long-term relations with key institutions and a balance between budget support and 'project-type' or regionally based interventions) are common to most of the studies by comparators, light or in depth, and echo concerns in DFID CPEs about too rapid a change of objectives and relationships and about the need to understand how government systems are actually delivering in the field. It seems doubtful that many more generic lessons would be derived by just commissioning more, or more ambitious, CPEs (as opposed to testing more specific hypotheses across a sample of country programmes).

   Recommendation: DFID management, in collaboration with Evaluation Department, should look at the case more cross-country evaluations with a strong lesson-learning component.

3. *Conclusion:* Donor practices in respect of reviewing country strategies have changed little despite the rhetoric of the Paris Declaration and the Accra Agenda for Action and despite the fact that objectively the amount of jointly-financed programmes and common strategies, hopefully increasingly owned and led by the host country, has risen significantly, and seems likely to continue to do so in many cases. As the Danish evaluation of aid to Uganda comments: "Progress on harmonisation and alignment has implications for future evaluations. Joint assistance strategies imply joint monitoring and evaluation, and it makes little sense to evaluate aid separately from the national strategy that it supports". Such country-led evaluations might look harder at the delivery mechanisms for aid and at its impacts and also play a role in supporting an evidence-based critique of individual and collective donor practices in the country concerned. If such an evaluation were available, the individual reviews of country strategy that donors will no doubt continue to need could be set against a properly evidenced and common frame of reference, and might assume the character of lighter reviews, though still independent of line management. It will be important that bodies such as IACDI do not discourage a more mutual process by their focus on improving accountability. In this connection, it is interesting that local involvement in Uganda was taken a stage further by the evaluation of the Poverty Eradication Action Plan (PEAP) published in 2008. This was an example

of a locally-led, even if donor-funded evaluation, set up to answer serious issues about how Uganda had approached the issue of poverty eradication at a time when a new planning framework was being put in place, and may point a direction of travel which could be developed.

Recommendation: DFID Evaluation Department and the Aid Effectiveness Team should consider proposing a deeper reflection, perhaps jointly by the DAC EVALNET and the DAC-hosted Working Party on Aid Effectiveness, on a new model of country-level evaluation which would have much stronger local input, and conceivably form an important element of processes for Mutual Accountability.

A.2 Thematic Evaluations

4. *Conclusion:* The volume of individual evaluations is in danger of outstripping the capacity of most stakeholders to assimilate lessons.

   Recommendation: DFID Evaluation Department should encourage a collectively financed and organised effort, perhaps under DAC/EVALNET or 3ie auspices, to select and publicise strong evaluation lessons in major areas of common interest on a regular basis in order both to improve lesson-learning and to reduce the temptation for each donor to evaluate the same topics.

5. *Conclusion:* The boundary between evaluation and research is worth further examination. Most of the comparator agencies have very modest research programmes, but DFID, which is a very large commissioner of research, is in a different position. It seems likely that there is scope also for greater collaboration (and greater use of evaluation findings) in thematic research areas, as is already developing on a small scale in DFID in the cross-cutting area of impact evaluation. This should take, as one starting point, that evaluations by other donors are likely to be as useful and of similar quality to those commissioned by DFID, and should always be included in the evidence to be examined, for example in portfolio reviews carried out by DFID's Investment Committee.

   Recommendation: DFID management should commission a study of how research and evaluation programmes might best work together to produce rigorous and evidenced lessons for policymakers.

6. *Conclusion:* It is evident from my survey that the most convincing evaluations are those which are based on adequate field work. This in turn links to the role of partner country institutions, whose role is normally essential to this. Several evaluations demonstrate that capacities of locally-based institutions and consultants are growing. It is time for a move to make greater use of genuinely joint evaluations with developing country institutions. Such an approach would require good quality control, but over time it would build capacity (particularly when institutions as opposed to ad hoc teams of consultants are the Southern partner), and it is evident that such genuinely joint evaluations would have more resonance at the developing country end: the evaluation system at present remains unduly donor-oriented.

Recommendation: DFID's Evaluation Department, like its comparators, should continue to encourage host country-led and genuinely joint host country/donor evaluation and use of local evaluation capacity (not just in a data collection role). It might experiment, as have some comparators, with local tendering of some evaluation work.

B. Quality of Management

On the second point, I interpreted the question as covering the whole evaluation cycle from selection of topic to management response to recommendations, including therefore not just Evaluation Department's own systems but also those of DFID more broadly.

7. *Conclusion:* DFID is reasonably placed in respect of its **operating environment** to encourage good quality central evaluations (on decentralised evaluations and reviews, see point 17 below) but could do even better by bringing in more staff with a background in research and evaluation, and by keeping evaluation managers in post for a four-year norm. It should be a high priority to put in place a plan to achieve these two linked objectives.

Recommendation: DFID management should ensure that Evaluation Department is able to develop a mix of staff with DFID operational experience and staff (including with no DFID experience) who have particular expertise in evaluation and/or research methodology.

Recommendation: DFID management should devise means to ensure that senior evaluation managers with DFID operational backgrounds stay on average for 4 years in post, without sacrificing career enhancement.

8. *Conclusion:* Specifically in terms of **resources,** my necessarily limited review suggests the following conclusions. DFID's budget for commissioning studies is a little above average for the group in relation to the size of the bilateral aid programme. Its complement of senior evaluators is however below average in relation to the size of the bilateral aid programme, and these evaluators are responsible for managing an above-average number of studies per year. DFID's evaluations are relatively numerous but also relatively inexpensive. DFID appears to lead joint studies more frequently than the average (though some agencies argue that it takes almost as many resources to participate in as to lead such an evaluation). In terms of resourcing, based on a necessarily limited comparison with the other agencies, DFID could consider either expanding the number of its senior evaluators or cutting the number of evaluations, perhaps at the same time increasing the resources for certain evaluations where this would significantly improve quality (but see Burt Perrin's report for the need to ensure value for money). It is encouraging that the DAC Evaluation Network will now be undertaking work on the systems and resources of its members, which may provide more comprehensive information in this area.

Recommendation: DFID Evaluation Department should report to IACDI on the results of the DAC Evaluation Network study, as a basis for further consideration of resourcing issues.

9. *Conclusion:* DFID's approach to **selection of evaluation topics** is in line with most comparators, though, as pointed out above, it is in the minority in its approach to country programme evaluations. It might however consider a closer link to the way that DFID objectives are set, supported by timely impact evaluation where relevant. (See the main report's description of the Dutch system.)

   Recommendation: DFID management should consider a system whereby all major objectives are covered by evaluation material on a multi-year cycle, thus giving greater ownership to the senior manager accountable for each objective.

10. *Conclusion:* DFID has an appropriate policy stance on **joint evaluation** with other donors, in line with most comparator agencies. This is consistent with the harmonisation dimension of the Paris Declaration.

    Recommendation: DFID Management and IACDI should build harmonisation into their requirements for accountability.

11. *Conclusion:* The alignment dimension of the Paris Declaration, and the vision in the Accra Agenda for Action of "country-led evaluations of policy performance" is far less developed. At present, there is no guidance to country offices on the case for **supporting the development of local evaluation systems**. This contrasts with important DFID initiatives in the area of better statistical systems. DFID, as a major provider of budget support, has a clear interest in tools that can drive up the quality of host government spending. Effective host country-led evaluation is one of these.

    Recommendation: DFID management should, as part of its focus on better evidence for policy-making, encourage country offices (e.g. via Directors' Delivery Plans) to include support for evaluation systems (and for impact evaluation) in their dialogue with host countries on the effectiveness of public spending.

12. *Conclusion:* DFID is in line with the essentials of good practice in respect of **quality controls**, on which the similarity of approaches among all agencies which operate by commissioning consultants (other than the EC, which has a more directive system) is striking. There are widely-shared concerns over variable quality of consultants. Dutch experience suggests that if evaluation agencies are to address this effectively, a prior step is to have experienced and qualified staff as argued for above.

    Recommendation: DFID Evaluation Department should consider the EC approach of limiting the number of evaluation questions in order to avoid reports that are too long and discursive.

13. *Conclusion:* I was asked to look specifically at DFID's practice of **commissioning consultants rather than carrying out evaluations with its own staff**. This is in line with the practices of Irish Aid (though Irish Aid is also trying an experiment of hiring an experienced consultant team leader with Irish officials with experience of the work area concerned as part of the team), NORAD, Sida and the European Commission. SADEV used to carry out studies using its own

staff, who mainly came from a research background. However, it has recently broadened its resource base and is moving towards a model similar to that of IOB/Netherlands. IOB/Netherlands have a mixed system where its senior evaluators ('Inspectors') lead all evaluations but routinely have consultants working as part of the team for each study. It is not practicable to make a judgement on the respective merits of each system on such a small sample of studies. DFID senior evaluators do seem to be at the upper end of productivity in terms of studies per senior evaluator (2.8 a year, compared to around 1.5 in NORAD and EC), which may limit their ability to manage the process as tightly as in these agencies. Again, securing the right balance in staffing in quality, permanence, and in terms of studies per senior evaluator, is probably the first issue to address.

Recommendation: For the present, the model of contracting our evaluations seems appropriate. Once DFID Evaluation Department has an appropriate staff complement with appropriate skills in place, it could consider experimenting with one or two evaluations led by DFID senior evaluators on the Dutch model.

14. *Conclusion:* DFID is in line with common practice on **disseminating reports**. However one weakness compared to former DFID practice is that there is, since the demise of the former Projects and Evaluation Committee, no senior DFID committee which automatically discusses the findings and recommendations of evaluations. It would be desirable that all centrally financed Evaluations should be discussed by a senior committee, perhaps the Country Programme Review Committee for Country Programme Evaluations and/or the Development Committee for thematic evaluations. By one means or another (perhaps an annual meeting) the Management Board itself should have an opportunity to consider key lessons learned from evaluations. If evaluations are not worth official time at least at the level of the senior committees of the organisation, why commission them?

Recommendation: DFID management should ensure that all evaluations are discussed by a senior departmental committee, and that the Management Board sees an overview of key lessons from evaluation once a year.

15. *Conclusion:* DFID's system for checking **whether action is taken on recommendations** has been improved from the point of view of monitoring with the introduction of checking by Internal Audit, and would now seem to be as robust as most of the comparators. However, as pointed out in Burt Perrin's report, this system is not a substitute for considered attention by senior management (see the previous recommendation). In addition, the NORAD and EC practice ('fiche contradictoire') of putting a summary of action taken into the public domain would be worth considering.

Recommendation: DFID should ensure that information on action taken on all evaluation reports is readily available on its website. It would be good practice in my view for all agencies to have an 'evaluation' button on their home page, leading directly to all published studies, (with EC-style quality scores),

management responses to each study, and a list of studies in process or planned – DFID should encourage the DAC Evaluation Network to set some general standards in this area (this is also relevant to the International Aid Transparency Initiative).

16. *Conclusion:* As for **outside perceptions** of the quality and relevance of evaluation reports, my enquiries suggested that DFID evaluation reports probably rate average (which means in practice quite low) in terms of generating Parliamentary interest, and below average for civil society, who at present pay much more attention to evaluations by multilateral agencies. It might be worth exploring further whether there should be a greater effort to interest CSOs. As noted in the report, the Dutch system in particular seems to generate a significant degree of Parliamentary interest, but this was not typical, except where reports were particularly critical.

    Recommendation: DFID's Information Department and Evaluation Department should examine what can be done to achieve greater public profile for evaluation findings. This should include discussions with civil society and other potential stakeholders.

17. *Conclusion:* DFID is poorly informed on what **evaluative work is being carried on outside Evaluation Department**, as is the case with several of its comparators. The decision by DFID management to advance pragmatically towards more end-of-project evaluation squares with concerns elsewhere not to reach for the tool of full 'evaluation' too readily. Studies by IOB/ Netherlands, UTV/Sida and EC show that sustaining quality of decentralised evaluations is not easy, and it will be necessary to allocate significant resources if the new system is to achieve its potential. (Several comparator agencies allocate 20-30% of evaluation unit time to assisting with decentralised evaluations.) It will also be necessary to consider how the expanded programme will meet the requirements of harmonisation and alignment indicated above.

    Recommendation: DFID management should determine how information on decentralised evaluation should be held in the organisation, and which unit should be responsible for ensuring that a proper database is available. Further discussion should take place on the proportion of Evaluation Department's resources to be devoted to supporting decentralised evaluation in DFID.

1. Introduction

1.1 My terms of reference required me to do three things:

i. Drawing on my institutional knowledge and a trawl of the relevant literature, to make comparisons between the quality of DFID's evaluations (based on a small number of between five and 10 of the sample of studies selected by Burt Perrin) and those of selected other agencies. The comparator agencies would be agreed with the review leader, Roger Riddell.

ii. Within the time constraints of the assignment, to make a preliminary assessment of DFID's Evaluation Department's (EvD) management systems and processes for ensuring the quality of evaluations undertaken and commissioned by DFID. This review would look at the process of commissioning evaluations, including EvD's policy of commissioning but not itself undertaking evaluations, and DFID's response to evaluations. This should, ideally, include discussions with key staff, consultants and, to the extent possible, comparisons with practice in comparable donor agencies.

iii. To produce a report of not more than 25 pages (excluding a two page Executive Summary and appendices) including comments on the assessments of the sample of evaluations, comparisons with other agencies, and recommendations for further strengthening of DFID's and especially EVD's role, in commissioning, managing, delivering and using high quality and effective evaluation studies, drawing on comparable agency practice and other innovations in evaluation methods (I fear that like many writers of evaluation reports, I have exceeded these limits).

1.2 This report is in two main parts: a comparison of six DFID studies (two Country Programme Evaluations and four thematic evaluations) with twelve studies by comparator agencies in similar areas; and a comparative assessment of EVD's (and, where relevant, DFID's) management systems and processes for ensuring the quality of evaluations.

1.3 It was agreed at the outset with Roger Riddell that the comparator agencies should be drawn in the main from the Nordic Plus Group of agencies. It was felt that these were better comparators than a more random choice of DAC agencies. It was also felt that I should concentrate on systems with reasonably straightforward institutional arrangements (unlike, say, Germany, with separate evaluation units in BMZ, KfW and GTZ). We therefore selected Ireland, Netherlands, Norway and Sweden as the bilateral comparators. Among these, all but Sweden have individual evaluation units, like DFID, and the Swedish model with a recently established fully independent evaluation agency (SADEV) alongside Sida's own evaluation unit (UTV) seemed worth

examination. We also concluded that the International Financial Institutions represented a model and a level of resourcing too far removed from DFID to be a good point of comparison[1] but that I should include the European Commission, a large multilateral donor with an evaluation unit of a size comparable to that of DFID.

[1] For example, the World Bank's Independent Evaluation Group produces 15-20 studies a year with a budget of $12 million, whereas DFID's Evaluation Department produces some 25 studies a year with an operational budget of £4.3 million. A DFID evaluation therefore costs a little under £175,000 on average, while a World Bank one costs some $700,000. For more data, see Annex 11.

2. A comparison of DFID evaluation studies with studies by comparator agencies

2.1 Introduction and selection

2.1 I should make clear at the outset of this section that it is impossible to make any valid statement about the comparative quality of DFID and comparator agency evaluations without a statistically appropriate sample, which would have to be far larger than envisaged in this study. No simplistic comparisons with Burt Perrin's findings in respect of DFID evaluations can be made because the choice of studies was not random, because, unlike the Perrin study, I merely read the evaluation reports themselves, and did not have access to the more extensive management data which he was able to use, and because I am not an evaluation expert. I agreed with Roger Riddell therefore that the approach would be to select comparator studies in areas as similar as possible to those DFID evaluations chosen from the sample examined by Burt Perrin with a view to establishing whether there were good practices in comparator studies that might be relevant to future DFID work.

2.2 It was agreed with Roger Riddell, after consulting Evaluation Department that the DFID evaluations for comparison would be:

- The country programme evaluations of Nepal and Zambia
- The thematic evaluations of Gender, HIV/AIDS, Private Sector Infrastructure and Voice and Accountability.

2.3 As noted above, I selected the comparative studies from the same general areas. The final list was:

- For country programme evaluations, EC-Tanzania, Ireland-Uganda, Norway-Zambia and Denmark-Uganda. I went outside the main comparator group in selecting the last of these as Netherlands and Sweden have no recent country programme evaluations, and as the Danish Uganda study is of some interest in its own right.
- For thematic evaluations, I selected the following:

  Private Sector: Country-led Joint Evaluation of the ORET/MILIEV Programme in China (China/Netherlands); Evaluation of Swedfund (SADEV (SADEV Report 2008:3);

  Voice and Accountability: 'Chatting and Playing Chess with Policymakers' (IOB/Netherlands); Evaluation of Democracy Support Programmes in Bolivia, Guatemala, Nicaragua and Peru (Sida);

  Gender: Promoting Women's Human Rights and Enhancing Gender Equality in Kenya (Sida); Evaluation of the 'Strategy for Women and Gender Equality in Development Cooperation (1997-2005)' (NORAD);

HIV/AIDS: 'Turning Policy into Practice : Sida's Implementation of the Swedish HIV/AIDS Strategy' (Sida); Evaluation of Norway's HIV/AIDS Responses (NORAD)

2.4 My summary scoring of these evaluations is in Annex 1 and the scores are aggregated in Annex 2, showing that I rated most studies overall as 'good' or 'acceptable'. These scores should however be read in the light of the caveats above. They are very much my own judgements, based solely on my reading of the reports, and do not purport to be any comment on the overall quality of the work of any of the evaluation units concerned.

2.5 This section discusses the country programme and the thematic evaluations in turn.

2. 2 Country Programme Evaluations

2.6 Country programmes are a prime 'unit of accountability' for almost all aid agencies. Significant proportions of total agency spending are delivered through country programmes, and country programmable aid is almost always[2] the largest proportion of bilateral aid disbursed at country level.

2.7 However, these facts do not necessarily translate into a strong case for country programme **evaluations**: in fact the case for such evaluations is quite contested. While most multilateral agencies (the World Bank, the Asian Development Bank, the UNDP and the EC being examples) conduct such evaluations to varying degrees of detail, it seems that by no means all bilateral agencies do likewise. In my sample of comparator agencies, for example, the Netherlands and Sweden, despite a Sida report of 2007 that came up with a proposal for trying out such evaluations, have no current intention to carry out country studies, and Norway appears to have conducted only one in recent years (the one on Zambia discussed below).

2.8 The reservations about single donor country studies derive from two main considerations: are they worth the cost, as opposed to lighter internal reviews, and do they make much sense when donors are increasingly co-financing an overall development programme and when the Accra Agenda for Action is (rightly) pressing donors to build accountability on local systems? The Sida study records that both Norway and the Netherlands discontinued earlier programmes of country evaluations on the first grounds and it underscores the problems inherent in trying to assess the impact of any single donor's aid. It also bemoans the fact that "there appears to be almost no discussion on *the partner countries' interest* in and use of country programme evaluations. The donor organisation's interests dominate the formulation and use". It suggests that "From the recipient countries' viewpoint, there ought to be at least three reasons for country programme evaluations:

[2] In a few cases, humanitarian aid may be larger: aid via centrally-managed schemes (e.g. for CSOs) is for almost all agencies and recipient countries smaller than country programmable aid, though it may still be significant (and is by no means routinely captured in country programme evaluations).

- What does the cooperation offer the country compared with other forms of additional resources?
- What can we report with regard to results to *our* citizens?
- Which donors are most effective and which are most useful to us?"

However, it admits that in practice there has been little interest in this on either side.

2.9 The Sida report offers a brief taxonomy of country programme evaluations:

1) "The most ambitious is to look at the *combined effects* of the country programme. This is naturally also the most difficult as it becomes necessary to determine which changes result from the individual country programme – parts of or the whole programme – and not from other factors(what evaluators call the *attribution problem).*

2) Another approach is to examine the *cooperation process,* i.e., the creation and implementation of the country strategy, while the individual components of the country programme are evaluated separately without assessing the overall result.

3) A third alternative is to only look at the composition of the strategy document and the country programme, in particular to appraise *assumptions, consistency* and *relevance,* and to try to draw conclusions on whether the contributions were well chosen in relation to the situation at the start of the strategy period and maybe also how it has changed in the country during the implementation period."

2.10 DFID's approach to the dilemma indicated above has been to opt for a relatively 'light-touch' evaluation of country strategies along the lines of 2) or maybe 3) above, an approach similar to that being taken by Irish Aid, for example in the case of Uganda. The EC has chosen a somewhat better resourced approach along the lines of 2) which makes possible some attempt to gather information on local impact (e.g. through focus groups) in addition to making use of documents, scoring systems and interviews of (usually) capital based and agency HQ based stakeholders. Norway and Denmark decided on more in-depth and long-period assessments of their contributions in Zambia and Uganda respectively, in line with some earlier Swedish studies in Botswana and Lesotho, and closer to the first option. Denmark however by no means evaluates all its country programmes[3], and in many cases uses an internal review process, as does Sweden, which does not aspire to the status of an independent evaluation. Most if not all agencies would consider the cost of the long-period Norwegian and Danish studies too heavy to be replicated across their main bilateral programmes. The Norwegian study was budgeted at NKr 2.7 million, or over £300,000. The World Bank's approach (which I have not investigated in detail) would appear to be a more strongly resourced version of EC practice.

[3] It has however carried out recent evaluations of Danish country programmes in Mozambique, Ghana and Benin.

2.11 The problem with DFID's approach (pointed out in some of my discussions with DFID users) is that the light-touch approach may be too light to produce evidence-based conclusions that go beyond conventional wisdom, given the propensity (documented by Burt Perrin) for DFID country teams to argue the toss in considerable detail. It was put to me that DFID's country programme evaluations were too modest in resourcing and content[4] in relation to the efforts now being made by DFID country offices to improve the analysis underlying their new country strategies. At the same time, senior management clearly values an independent assessment of the effectiveness or otherwise of country strategies, particularly as new strategies are produced[5]. One suggested to me that a short visit by an experienced person with no axe to grind (as in the DAC Peer Review process) could produce in a few days as useful a product as a team of consultants on present lines[6].

2.12 In this connection, the **Irish evaluation of its country strategy in Uganda** is a product very comparable (indeed using the same firm of consultants) to the DFID model, even if more lightly resourced (a budget of 95 consultant days) and dealing with a relatively small programme in absolute terms[7]. Good practices evident in this report include **a direct link to Irish Aid's consideration of its next country strategy for Uganda** (something not always achieved in DFID country programme evaluations, partly because of 'bunching' of consideration of new strategies); **incorporating field visits** even in a tight programme in country; and conducting a **staff perceptions survey** of the Irish Aid office in Kampala (which revealed that 83% of respondents felt that they needed more time to visit the field and that 64% needed more time to analyse and discuss strategy). The conclusions appear to follow logically from the (limited) evidence available, and to be well attuned to the strategy review process. The evaluation, like those of DFID, is almost exclusively a donor-oriented product.

2.13 The **European Commission evaluation of its country programme in Tanzania** was also commissioned to coincide with consideration of the EC's country strategy, and apparently had considerable impact on the new strategy. It was more strongly resourced, involving 10 named consultants, a majority of whom, along with the co-leader, were Tanzanian, and a budget of Euros 200,000. The arrangements for accessing local knowledge included a panel of Tanzanian experts, and focus group discussions on eight specific field activities considered to be representative.

[4] Interview with FCPD, 4 March.
[5] Interview with a DFID Director-General, 23 March.
[6] Interview with a DFID Regional Director, 27 April.
[7] Irish Aid has been conducting Country Programme Evaluations since 2002. Until recently, their Country Strategy Papers were of three year duration and the policy was to evaluate them at least every second time, i.e. every 6 years. Irish Aid's CSPs are now more often of longer duration, up to 5 years, to align with government PRSPs. These are evaluated in their final year - leading into the development of the next CSP.

2.14 Interesting features of this evaluation, which was among very few given a 'very good' rating under the EC's quality assurance system, which is described in Section 3 below, included:

- inviting the consultant to propose and clear with a reference group **no more than ten evaluation questions** which the study should address, each with an explicit criterion or criteria for judging whether the answer to each question was positive or negative
- defining **an intervention logic** around how the EC programme might contribute to poverty reduction
- using the **field studies and the expert panel to help 'triangulate'** emerging findings
- **frankness on areas where the evidence to support conclusions was relatively weak** (e.g. the conclusion that "General Budget Support has been found to have an equal or even better potential [than other forms of aid] to contribute to poverty reduction" is said to be "connected to evidence through a long, and relatively weak, chain of reasoning").

2.15 Most of these good features do not require resources so much as willingness by the consultants, encouraged by suitable terms of reference and consultation, to focus and to think things through.

2.16 The EC study, for all its use of local expertise and the inclusion (standard for EC country programme evaluations) of the country's Brussels Ambassador in the reference group, still has very much the character of an evaluation aimed almost exclusively at the donor. Accountability to the Commission, the European Parliament and the EU public, are writ large up front, even though there are also some useful lessons with some scope for generalisation.

2.17 In line with EC requirements, the evaluation tries to be explicit about impact of the EC's aid, concluding:

"The EC has contributed to this progress in many ways. First, the general budget support has been additional in terms of expenditures in the priority sectors, and therefore in the results achieved. Second, remarkable contributions have been achieved through a mix of sector policy dialogue and large financial support in sectors like education and roads. Some projects have also achieved impressive impacts, although at a lower scale, for instance a very substantial increase in farmers' income in some tea/coffee growing areas.

This evaluation has paid special attention to the impact of EC-supported road investments on poverty reduction. An in-depth field study has identified impressive impacts of road investments in terms of poverty reduction, especially through "equitable access to basic services". Stand-alone road projects are, however, insufficient for achieving such impacts, which stem only from a well maintained road network, consistently connecting trunk roads, feeder roads and local roads."

2.18 In terms of IACDI's and DFID's requirements, it might be interesting to experiment with a couple of somewhat more generously resourced country programme evaluations of this kind, linked to DFID's strategy review process. However, this should not obscure the need for some more radical thinking, as set out at the end of this section.

2.18 Both the **Nordic studies** are of an altogether more ambitious character. Each study attempts to document impact of the donor's aid over more than a decade, and involved significant fieldwork and a perceptions survey as well as the customary review of documents and interviews with key donor and recipient stakeholders.

2.19 The **Norway-Zambia evaluation** covers the period 1991-2007 and is marked by its adoption of the political economy model of a 'neo-patrimonial state' as its core hypothesis about why income per head in Zambia was still, at the time of the evaluation, lower than at Independence. This is a perhaps an oversimplified view (one might think that the secular decline in copper prices was also a major factor), but the DFID country programme evaluation of Zambia also speaks of 'entrenched patronage politics' as a major concern. Defending its thesis, the report notes that "We find it significant that, in presentation of earlier versions of this report, very few Zambians disagreed with the analysis of Zambia as a neo-patrimonial state (though none of those consulted were actually members of government). The analysis was more contested by non-Zambians."

2.20 The evaluation, despite terms of reference running to 10 pages, has at its heart just three questions:

- How do power relations within Zambia and amongst the donor community affect the scope for achieving pro-poor outcomes through development cooperation?

- How do power structures at different levels of Zambian society influence the potential of aid to reduce poverty?

- How do domestic and international power relations in Zambia affect the outputs of Norwegian financed pro-poor interventions?

2.21 Even though there are a large number of sub-questions (arguably too many), this simple list gives focus and structure to the report.

2.22 The evaluation argues that the 'neo-patrimonial' model fits the Zambian political scene well, although it detects signs of change to a more modern political society. It takes the donor community, including Norway, to task for 'de-politicising' inherently political issues, and focusing too much on macroeconomic issues. It highlights concerns that harmonisation may be limiting the space for third parties (e.g. Civil Society) to be heard in what is seen as increasingly a dialogue between the donor group and the government. And it questions the effectiveness and sustainability of the specific Norwegian interventions which it addressed in detail. (The executive summary is reproduced as Annex 3 for ease of reference.)

2.23 On the positive side, the evaluation raises serious issues that are too often handled 'en passant', as in the DFID Zambia CPE, and which have resonance far beyond Zambia and Norway, thus perhaps justifying its cost. It gets away from what might be considered too much of a macroeconomic "straightjacket" that has not always been useful in answering the more fundamental questions about aid, governance and dependency. On the negative side, the report to some extent reads as a prisoner of its intellectual model and its own political analysis is very broad-brush. The report had an unusual degree of profile: the Government of Zambia (although members of the Government were consulted) formally disassociated itself from the report, while in Norway the Opposition made considerable use of it to criticise the government's approach to aiding African governments.

2.24 The **Danish study of aid to Uganda** is similarly ambitious, covering the period 1987-2005, but approaches its task rather differently. In the first place, it was set up as a joint evaluation with the Office of the Prime Minister of Uganda, and though in practice it seems to have been very expatriate-led and with strong Danish ownership, this does give it a different cast. For example, there were two reference groups, one chaired by DANIDA in Copenhagen and the other chaired by the Office of the Prime Minister in Kampala. It is, from this point of view, very striking that the GoU were prepared to put their name to a report with such judgements as:

- *"There is a widespread belief, however, that the UPDF deliberately provoked insecurity as a cover for Uganda's involvement in the illegal exploitation and export of minerals and other resources from the area [NE Congo]. The army's privileged position in Ugandan society has allowed army officers involved in these and other corrupt practices to go unpunished".*

- *"By the end of the period under review, much of the early enthusiasm for the NRM government had been eroded, with both public and donors concerned that Uganda has become a* ***de facto*** *one-party state and with highly personalised presidential power. Associated with the Movement's monopoly on power, Uganda has come to appear like other patrimonial states with high levels of corruption and the use of public resources to buy support. The creation of new Districts in areas supportive of the Presidency is a conspicuous example of this form of patronage."*

2.25 Another point of interest is that, consistent with the trend of DAC discussion at the time, the report has a 'layered' approach, starting with Uganda's progress, then assessing the overall role of aid, and only then addressing the success or otherwise of Danish interventions. Its conclusions on the role of aid are worth quoting:

"There is a radical critique which argues that aid inevitably undermines the domestic accountability of regimes and enables them to side-step the challenge of raising resources domestically. This is a legitimate concern, but it is not plausible that Uganda could have achieved similar levels of investment, growth, expanded public services and poverty reduction without the aid that it received. Our judgement is that aid in aggregate made a clearly positive economic, social and political contribution. This forms the context for efforts to evaluate the contribution of Danish aid in particular."

2.26 This is probably the nearest that any of the country programme evaluations reviewed comes to examining a counterfactual.

2.27 The report is underpinned by 13 separate studies of different areas of intervention by DANIDA, and it seems to have been able to be appropriately critical of some early, high-profile, projects (grain stores, milk plant) that proved inappropriate and unsustainable. Overall, the evaluation shows that with adequate resources (in this case 12 person-months of international staff time and 8 person-months of local staff time) robust and at times hard-hitting conclusions can be reached. The bigger question is the extent to which the report had real influence either in Denmark or more broadly to justify the investment.

Conclusions: Country Programme Evaluations

2.28 This review of approaches by the comparators leads to three main conclusions:

1. While all donor agencies need for **accountability purposes and for future planning** to have some way of assessing at least the effectiveness of their country strategies (if not the impact of their country programmes), the way in which they do so is quite varied, ranging from internal reviews to light evaluations as practised by DFID, more strongly resourced ones such as those of the EC (or the World Bank), or the occasional 'long-period' study as tried at different times by Sida, NORAD and DANIDA. **It is not clear that one approach is inherently superior to another,** though both DFID and EC evaluators would argue that their models provide consistent and independent approaches which give relatively high coverage of the bilateral programme**.** Donors need to consider how their accountability and assessment needs can be met in a cost-effective but sufficiently robust manner. While DFID could experiment with one or two better-resourced CPEs, the benefits of more complex and expensive evaluations may not offset the cost and time incurred. The quality of data and the adequacy of frameworks for monitoring and evaluating country programmes are significant issues for comparator agencies as well as for DFID.

2. Many of the **lessons** (for example about maintaining long-term relations with key institutions and a balance between budget support and 'project-type' or regionally based interventions) are common to most of these studies, light or in depth, and echo concerns in DFID CPEs about too rapid a change of objectives and relationships and about the need to understand how government systems are actually delivering in the field. **It seems doubtful that many more generic lessons would be derived by just commissioning more, or more ambitious, CPEs (as opposed to testing more specific hypotheses across a sample of country programmes).**

3. Donor practices in respect of reviewing country strategies have changed little despite the rhetoric of the Paris Declaration and the Accra Agenda for Action and the fact that objectively the amount of jointly-financed programmes and common strategies, hopefully increasingly owned and led

by the host country, has risen significantly, and seems likely to continue to do so in many cases. As the Danish evaluation of aid to Uganda comments: "Progress on harmonisation and alignment has implications for future evaluations. Joint assistance strategies imply joint monitoring and evaluation, and it makes little sense to evaluate aid separately from the national strategy that it supports". **It would be worth a deeper reflection, perhaps jointly by the DAC EVALNET and the DAC hosted Working Party on Aid effectiveness, to try a new model which would have much stronger local input, and conceivably form an important element of processes for Mutual Accountability**. Such country-led evaluations might look harder at the delivery mechanisms for aid and at its impacts, and also play a role in supporting an evidence based critique of individual and collective donor practices in the country concerned. If such an evaluation were available, the individual reviews of country strategy that donors will no doubt continue to need could be set against a properly evidenced and common frame of reference and might assume the character of lighter reviews, though still of independent of line management. It will be important that bodies such as IACDI do not discourage a more mutual process by their focus on improving accountability. In this connection, it is interesting that local involvement in Uganda was taken a stage further by the evaluation of the Poverty Eradication Action Plan (PEAP) published in 2008. This was an example of a locally led, even if donor funded evaluation, set up to answer serious issues about how Uganda had approached the issue of poverty eradication at a time when a new planning framework was being put in place, and may point a direction of travel which could be developed.

2.3 Thematic Evaluations

2.29 This section compares the DFID studies with a couple of studies by other agencies in the same general fields with a view to learning lessons. These are set out in respect of each field, with a few common issues highlighted in conclusion.

Private Sector

2.30 I compared DFID's desk review of private sector infrastructure facilities with an evaluation of one institution, Sweden's Development Finance Institution, Swedfund, carried out by SADEV and one programme, the Netherlands' mixed credit programme (ORET/MILIEV) in China, which was evaluated by the Chinese National Center for Science and Technology Evaluation (NCSTE) in partnership with the Dutch evaluation unit, IOB – an unusual example of a genuinely partner-country led evaluation.

2.31 While each evaluation had its particularities, I was struck by two common themes. The less important, but operationally very significant one, was that information was very poor, especially on impact. The DFID desk review strongly criticised data availability, and the management response did encouragingly address this point. The Swedfund evaluation was even more poorly served, to the point where the **only** robust conclusion was that M&E

needed to be radically improved. The China evaluation also suffered from the fact that a transfer of responsibility between Chinese Ministries resulted in loss of all the back files – though in this case a combination of information available in The Hague and extensive field surveys provided a solid data base.

2.32 The reasons for this problem, while different in detail, have to do with the problems inherent in dealing with the private sector: for example, once Swedfund had sold out of a company there was no reason for the company to continue to supply data. Information on impact and sustainability is always therefore likely to be a problem with facilities like those supported by DFID or the Netherlands or by Development Finance Institutions like Swedfund.

2.33 The more important one, to my mind, is that none of the three studies really got to grips with the counterfactuals and the fact that spending has an opportunity cost. Thus:

- DFID has invested £185 million in the various facilities that it has supported (and of which it has often been the lead supporter). For this sort of resource, how does its return compare with alternative ways of leveraging private investment?

- Swedfund has a portfolio of about SEK 1 billion (approx. £100 million). How far has that delivered genuinely additional investments? The evaluation does at least attempt to address the question, but its answer is far from convincing: "All in all of the cases we have looked at we are of the opinion that Swedfund's additionality has been high in 12, medium in 4 and low in 13 of the investments. In 18 of the cases we do not feel that we have enough information to form an opinion on the issue". It also notes that of 12 firms that replied to its questionnaire, 8 considered that they would have invested anyway, with or without Swedfund backing. It quotes IFC staff as questioning the cost-effectiveness of small national DFIs. These pointers to the need for possibly fundamental reappraisal of the real value-added of Swedfund are not however pursued in the conclusions of the evaluation.

- The Dutch mixed credit programme to China appears from the evaluation to have good or less good results very much according to the strength of the Chinese partner and the quality of Dutch equipment. However, Dutch as well as Chinese participants agree that Dutch equipment is supplied at above world market prices, and the evaluation concludes that the links created by the programme "are not commercially sustainable, and keep requiring Netherlands subsidies and Chinese tax exemptions". This finding gives cause to doubt the whole rationale for the programme, but the conclusion reached is merely that there is a case for limiting the number of times a Dutch firm can use the facility in the same country.

2.34 Thus in each case, there seems to have been some reluctance in these private sector oriented evaluations (sometimes for lack of clear evidence) to tackle what could be major issues.

2.35 In terms of evaluation practice, the most interesting study is the Chinese one. Here, the leading role of the NCSTE translated into an impressive programme of fieldwork covering 35 projects. Four separate teams were involved, all using

a standard check-list. No fewer than 17 provincial level and 2 national level stakeholder workshops were held. All this provided a robust amount of 'real-world' information that gives the study's findings on, for example, problems of management, pricing and technology transfer solid foundations. It is also evident that the role of NCSTE facilitated ownership by the key Ministry in Beijing. The Dutch IOB seems to have played a very constructive role in guiding the methodology and in transferring concepts, including through their Joint Coordinator. The NCSTE believe that it has visibly improved its professional capacity during the exercise. The presentation of the final report is excellent. Both the DFID and the SADEV studies lose impact from the lack of this field-level underpinning, as well as from the deficiencies of central M&E systems.

Voice and Accountability

2.36 Here I compared the joint donor study, managed by DFID, on Citizens' Voice and Accountability, with a Sida study of democracy promotion through NGOs in Bolivia, Guatemala, Nicaragua and Peru and a Dutch evaluation of grants to campaigning NGOs, engagingly titled "Chatting and Playing Chess with Policymakers".

2.37 These studies have a common theme of working with civil society organisations to improve the lot of poor people and to improve the conditions that limit their scope for voice and self-reliance but adopt very different methods. The DFID-led joint study was extremely complex, involving several distinct phases, country case studies commissioned by different donors and a heavyweight synthesis report. The Sida study was also ambitious, involving four country case studies of activities by a single NGO (Diakonia) to promote leadership in organisations representing marginalised groups in Latin America. Unusually, one motive for undertaking the study was methodological: a wish to test a method known as 'outcome mapping', developed at IDRC, and which required three interviews with (ideally) the same people over an 18 month period. In practice, this design had to be adapted to the fact that people move around. The Dutch study was a simpler affair, largely based in practice on six case studies of attempts by Dutch-based NGOs to influence Dutch companies and the Dutch government on international economic issues, such as the future of the EC sugar regime.

2.38 Each study has its strengths and weaknesses. The DFID study can be criticised as very expensive and time-consuming for what it delivers, and it is not clear how much impact on policy it will have. Nevertheless, it presents interesting findings which provide a basis for further discussion of how donors might address what is on any reckoning a complex area. The Swedish study produced an impressive amount of (largely qualitative) data, which enabled the final report to raise many challenging issues about the attempt to develop leadership of organisations devoted to raising the voice of the poor and marginalised. Although these are not always easy to read, and although there are no clear recommendations, the study is thought-provoking and sheds light on an

important aspect of the same theme. The Dutch study succeeds in developing a basis to assess the efficiency and effectiveness of advocacy and lobbying, and responded to clear interest from the Dutch Parliament; but like the Swedish study it raises as many dilemmas as clear actionable recommendations.

2.39 The main value of all three reports is in providing a structured approach to assessing how donors (and all three might be regarded as rather donor-centric, despite the importance of the field work of the DFID and Sida studies) can and should operate with civil society in its various manifestations in what are likely to be politically contested circumstances, whether in developing countries or in donor countries. In all three cases, Northern CSOs were represented in the reference groups, and the studies seem likely to have had some impact also on their consideration of how to position themselves on issues of political change and effective advocacy. All three also demonstrate the value of real field-based (or in the Dutch case, campaign-based) work as a starting point for learning lessons. The Dutch study graphically illustrated the limits of desk studies in relation to these complex areas.

Gender

2.40 On gender I compared the DFID gender evaluation with a quite similar NORAD evaluation of Norway's experience with gender mainstreaming, and with a very different piece of work – a decentralised evaluation commissioned by Sida's African Department of a grant to UNIFEM for work on women's empowerment in Kenya.

2.41 The NORAD study considers both central activity and country case studies in Afghanistan, Bangladesh and Zambia – as will be seen below under HIV/AIDS, a fairly classic approach to looking at the effectiveness or otherwise of a strategy. It was designed however with a strong emphasis on lesson-learning for the future, rather than just an assessment of action to date. The evaluation faced familiar problems of inadequate data and record-keeping, but was able to come to what would appear to be well-grounded conclusions.

2.42 These in many ways parallel those of the DFID evaluation. Few resources were allocated centrally: indeed the 'gender fund' was abolished in 1999, and the full-time adviser post was replaced in 2001 by part-time positions. The evaluation documents the gap between rhetorical statements to Parliament and the lack of reference to gender considerations in the annual Ministry Instructions to NORAD. The Afghanistan Strategy, rather astonishingly, paid no attention to gender.

2.43 There is an interesting discussion of the impact of new aid modalities, which are seen as in practice tending to marginalise gender aspects, while nevertheless having some potential for better 'engendering' national strategies and budgets – opportunities that Norwegian country offices were not resourced to seize. The evaluation therefore supports the concept of a time-limited fund to encourage work around such opportunities – an idea that was indeed endorsed by the Minister shortly after the evaluation. Overall, I felt that the evaluation was very similar to the DFID one, but I gave it a slightly higher grading than Burt Perrin did for the DFID study on the grounds of what seemed to me a better link between evidence and conclusions.

2.44 The Sida study by contrast was a relatively simple assessment, done in the space of a month or so by three local consultants, with a view to assisting the African Department decide on a new phase of a contribution to certain aspects of UNIFEM's work in Kenya. I chose it as an example of what must be a large number of such studies, timed to feed into a decision on whether to continue financing the activity.

2.45 To my mind, the study was a 'fit-for-purpose' response to a necessarily limited brief. In the time available, the team managed a wide range of interviews with key staff in UNIFEM, the Government of Kenya, Sida and other bilateral donors (CIDA being a co-funder of aspects of UNIFEM's programme), and the main CSO recipients, including field visits to several different Districts of Kenya. Good use was made of case studies (suitably anonymised except where individuals had given consent to be named). A particularly telling one was that of a Chief who gave the following lessons learned from the training he had experienced under the project:

- "That the children belonging to families where widows are not allowed to inherit (matrimonial property) undergo a lot of suffering.
- When widows are empowered they take better care of the home than when the husband was alive.
- Sometimes when the husband dies the woman starts to gain weight and the children start attending school. The woman had been oppressed.
- Women must be empowered. Even for us still alive we should give our wives part of our salary. My children call my wife Minister for Economic Planning. Whenever I earn she carries the family basket (budget) and she is able to run the home well. When I try to run it the money gets finished very quickly.
- In the Church I belong to a group of men who are saved and are referred to as *'those that the wife has sat on.'* Such men have dialogue with their wives and are doing well economically. You see the church encourages equality among men women".

2.46 The consultants' recommendations were practical (e.g. on programme management), and they were ready to look quite broadly at the issues – for example suggesting that consideration be given to why Tanzania and Uganda had made more progress than Kenya in some aspects of human rights and gender. They correctly highlighted serious issues around sustainability – not just for the CSOs but also for the gender unit of the relevant Kenyan Ministry, whose operational costs appeared to be wholly dependent on UNIFEM – though they could perhaps have said more about how to tackle this issue. They were also perhaps rather narrow in tackling the difficulty of identifying Sida's particular contribution to UNIFEM by proposing additional monitoring arrangements rather than by a more radical merging of Sida and CIDA finance. Overall, however, this was a practical and proportionate piece of work for those considering the future of the project.

HIV/AIDS

2.47 On HIV/AIDS, I compared DFID's major formative evaluation of the UK Government's strategy on HIV/AIDS in developing countries with two Nordic evaluations with a similar theme – an evaluation by Sida in 2005 of Sweden's own strategy towards HIV/AIDS in developing countries and a 2008 NORAD evaluation of Norway's responses to the pandemic.

2.48 While there are many points of similarity between the three evaluations (for example the use of country studies as a main input), the differences are illuminating (and not just the issue of cost, both Nordic ones - 55 consultant-weeks for Sida and 60 for NORAD - being much less expensive than the £750,000 DFID study).

2.49 The Sida evaluation is perhaps the more interesting of the two Nordic ones in that, as its title ('Turning Policy into Practice' implies) it takes a serious look at how strategies may be expected to influence institutional behaviour.

2.50 The Sida evaluation is presented as a formative evaluation (as with DFID) and with a clear purpose - "generating knowledge and creating discussions regarding Sida's co-operation in the area of HIV/AIDS to be used in Sida's future work. The evaluation process is intended to feed into the planned scaling-up of the HIV/AIDS-related work of Sida departments and embassies" - even though, in this case, the strategy in question had already been in operation for about 5 years. However, unlike the DFID case it had the advantage that Sida had carried out a baseline study in 2000/01 (the only example of a baseline study of a strategy which I found, and surely something that DFID and others should emulate), and that a more recent desk study of 21 major projects had already tested the extent to which mainstreaming was a reality. Another good practice, to my mind, was the decision to limit the scope of the evaluation to the question of how Sida had responded to the strategy, rather than going into the much more problematic issue of possible impact on the ground. While this gives the evaluation a very donor-oriented aspect, it does enable the study to focus.

2.51 A particularly interesting aspect of the study was the production, early on, of a 'conceptual paper' about how a strategy might be expected in practice to influence institutional behaviour. This drew on work by a North American author, M. Bemelmans-Videc, entitled "Carrots, Sticks and Sermons. Policy instruments and their evaluation". The thesis of this work is that policies can have their effect through three kinds of incentive: 'carrots' (e.g. more resources), 'sticks' (e.g. criteria for evaluating job performance) and 'sermons' (e.g. persistent top management expressions of interest, policy papers etc). The categorisation proved apt to the Swedish case, where the evaluators conclude that initially 'carrots' had proved somewhat ineffective on their own, and that it was only when 'sermons' became more consistent and a few 'sticks' were put in place (e.g. in the shape of Sida's annual letter of instruction from the government) that institutional support for the 1999 strategy really took off. So the Sida evaluation has a good deal of interest for all concerned in institutional change, and gives a dimension lacking in the DFID and NORAD evaluations of HIV/AIDS (and indeed of gender mainstreaming).

2.52 The NORAD evaluation is based on what I considered to be perhaps the best Terms of Reference of all the evaluations that I read. They were admirably clear in what was required, highlighting just five issues, on each of which they set out what approach should be taken:

- "Assess progress towards key outcomes related to the national HIV/AIDS response
- Assess the factors affecting the outcomes (substantive influences)
- Assess key Norwegian contributions (outputs) to outcomes
- Assess the Norwegian partnership strategies
- Extract lessons learnt, findings and recommendations on how to enhance the development
effectiveness of the Norwegian HIV/AIDS response at the country level."

2.53 Like the Sida evaluation, the TORs do not expect the consultants to do everything: thus, they invite them to specify which DAC criteria they consider to be the most relevant at country level.

2.54 The timing was however somewhat more problematic, in that the Norway's HIV/AIDS strategy was very recent. The TORs attempt to resolve this by stating that:

"The primary use of the evaluation is not to provide input into a revision of the Norwegian HIV/AIDS policy since the current Position Paper was introduced in November 2006. The evaluation should provide strategic guidance towards the implementation of the Position Paper, and in particular strategic guidance for enhancing the development effectiveness of the Norwegian HIV/AIDS response at the country level. The Position Paper could be adjusted based on inputs from the evaluation". They therefore suggest that "The main purpose of the evaluation is to ascertain results (accountability), fill knowledge gaps, provide lessons learnt and suggest recommendations on how to enhance the development effectiveness of the Norwegian HIV/AIDS response at the country level." This is a more ambitious brief than the Swedish evaluation, and more in line with that of DFID.

2.55 The evaluators did their best to catalogue the progress achieved in each of the countries for study (Ethiopia, Malawi and Tanzania), in the process, reinforcing the fact that good information on trends remains very scarce, even though the general message was broadly positive. Norway's efforts were categorised, the evaluators identifying Norway as a very responsive and pragmatic donor, to the extent that some greater strategic focus might be desirable (including in Norway's funding of NGOs). Norway's extensive international efforts were also assessed positively.

2.56 Where the evaluation was to my mind less successful was in detailing the sort of recommendations that the TORs called for ("Which strategic focus and actions will have the greatest developmental "pay-off"? What corrective actions are recommended for new, ongoing or future Norwegian work on the outcome in the country? What are the main lessons that can be drawn from the outcome experience that may have generic application? What are best practices

in designing, undertaking, monitoring and evaluating outputs, activities and partnerships around the outcomes?"). In fact, the recommendations appeared very modest considering the volume of evidence gathered in the evaluation. The DFID evaluation appeared to have more focussed recommendations in this respect, while missing the institutional change message of the Sida evaluation. My final thought, here as in the gender case, is that the high degree of commonality of questions, but differences of experience and approach, between the evaluations is evidence of the value of synthesis studies such as that carried out in this area by DANIDA, so that as many lessons as possible are drawn from the now quite impressive body of evidence.

Conclusions: Thematic Evaluations

2.57 A first conclusion of this brief review is that, taken together, there is a very large body of knowledge on many important development themes present in evaluations. **Indeed it is impossible to look, as I have done, at even half a dozen reasonably like-minded agencies and not wonder whether collectively the volume of individual evaluations has not outstripped the capacity of most stakeholders to assimilate lessons.** As the DANIDA synthesis of HIV/AIDS evaluations just referred to put it: "Organisations should consider harmonising and aligning their evaluation activities in order to utilise the resources used for evaluations better in terms of addressing the questions that globally needs answering".

2.58 The very useful DAC database (the Development Evaluation Resource Centre, or DEReC) lists for example 67 studies on gender and by coincidence the Danish survey of evaluations on HIV/AIDS also finds 67 such evaluations. Much of this literature has little resonance outside the agency that commissions it (as a specific example, two Dutch impact evaluations of basic education in Uganda and Zambia, both with interesting lessons (for example in the Uganda case on the importance of better school management) were not referenced in DFID's Portfolio Review of aid to basic education). Lesson learning could be enhanced by a more systematic approach to synthesising the results of these numerous evaluation studies on similar themes, as in a NORAD review of gender evaluations and the DANIDA review of evaluations of HIV/AIDS. At present, such syntheses appear to be carried out 'as and when' by individual agencies. **A collectively financed and organised effort, perhaps under EVALNET or 3ie auspices, to select and publicise strong evaluation lessons in major areas of common interest on a regular basis might both improve lesson-learning and reduce the temptation for each donor to evaluate the same topics.** Joint evaluations, for all their transaction costs, do have the virtue of having simultaneous resonance in all the commissioning agencies. The Joint Evaluation of General Budget Support shows that this can have something of a 'critical mass' effect in the right circumstances.

2.59 Secondly, the thematic evaluations raise interesting questions about the boundary between evaluation and research. Evaluations such as the one commissioned by Sida of Diakonia programmes in Latin America have an explicit link to anthropological studies, and its carefully-written and sometimes quite nuanced conclusions are what one might expect from anthropological

research. Most of the comparator agencies have very modest research programmes, but DFID, which is a very large commissioner of research, is in a different position. **DFID could usefully commission a study of how research and evaluation programmes might best work together to produce rigorous and evidenced lessons for policymakers.** There has already been some collaboration, for example on impact evaluation (joint funding of 3ie, working together on influencing the World Bank human development impact evaluation programme, Evaluation Department advising on quality of randomised trial on HIV interventions in Africa funded by the Central Research Department), where there is real scope for jointly promoting high quality research, and where there is plenty of potential for working together with the ESRC on impact evaluation funding of research. But there must surely be scope also for greater collaboration (and greater use of evaluation findings) in thematic as well as cross-cutting research areas. As noted above in the case of education, this should take as one starting point that evaluations by other donors are likely to be as useful and of similar quality to those commissioned by DFID, and should always be included in the evidence to be examined.

2.60 A third conclusion is that while desk reviews have their place (and, as suggested above, more syntheses of what already exists would be desirable), **the most convincing evaluations are those which are based on adequate field work**. This in turn links to the role of partner country institutions, whose role is normally essential to this. Although, as suggested above, the joint China/Netherlands evaluation of a mixed credit programme failed to raise some important issues, it certainly showed that extensive field work could be carried out and effectively analysed by an indigenous research organisation. Other studies (such as the EC/Tanzania and Sida gender evaluations mentioned above or the Norwegian study of CSOs in Northern Uganda mentioned below) demonstrate that capacities of locally-based consultants are growing also in Africa. This suggests that a move should be made to make greater use of genuinely joint evaluations, as in the China/Netherlands case. Such an approach would require good quality control, but over time it would build capacity (particularly when institutions as opposed to ad hoc teams of consultants are the Southern partner), and it is evident that such genuinely joint evaluations would have more resonance at the developing country end: the evaluation system at present remains unduly donor-oriented. Too many of the evaluations covered above were very donor-oriented. This needs to change.

3) A comparative assessment of Evaluation Department's management systems and processes for ensuring the quality of evaluations undertaken and commissioned by DFID, taking into account the practices of comparator agencies

3.1 Methodology and approach

3.1 For this part of the study, I have looked through the 'project cycle' for studies commissioned by DFID's Evaluation Department, and compared it to the practices of comparator agencies (Irish Aid, IOB/Netherlands, NORAD, SADEV, Sida and EuropeAid). I have used the questionnaire employed by the Peer Review and Evaluation Division of the OECD-DAC as a means of systematising the data, but have somewhat re-cast the DAC template in order to concentrate on the specific issues covered in this study. (As noted above, I have NOT drawn any conclusions about the quality of evaluation management systems from my survey of a very limited number of actual evaluations.)

3.2 This revised structure asks seven questions, as follows:

1. *Is the operating environment favourable to high-quality evaluations? This covers DAC questions about the policy and institutional setting, and about financial and staff resources [DAC Questionnaire Sections 1, Policy; 2, Impartiality, Transparency and Independence; and 3.1-3.4, Resources and Staff]*

2. *Is the way in which the topics for evaluations are decided likely to encourage relevance and buy-in? [DAC Questionnaire 6.1-6.3, planning and priority setting; 6.8, approach to multilateral aid]*

3. *Is the approach to involving partners and other donors in evaluations logical and reasoned? [DAC Questionnaire 4.1-4.3, involvement of partner countries; 6.4-6.7 involvement of other donors]*

4. *Are evaluations managed with a view to a quality product? [DAC Questionnaire 3.5, selection of consultants; 5.1-5.3 managing for quality]*

5. *Are findings disseminated in a way likely to secure impact? [DAC Questionnaire 7.1, 7.3 and 8.4]*

6. *Are recommendations followed up systematically? [DAC Questionnaire 7.2 and 8.3]*

7. *How are evaluation products perceived by outside stakeholders? [DAC Questionnaire 4.5, local stakeholders; 5.4 and 7.4 development agency staff; 8.1-8.2 other stakeholders, including Parliament and public]*

In addition, I have looked at an eighth question, not included in the DAC Questionnaire, namely:

8. What role does the evaluation department play in relation to evaluative material commissioned by other parts of the agency?

3.3 Annex 4 contains DFID's written response to the questionnaire, which may be compared with the more summary responses that I have drafted on the basis of my own discussions with the comparator agencies, in Annexes 5-10. These attempt to offer a brief but structured view of each system, and were cleared in draft with the comparator evaluation unit concerned (to all of whom I am extremely grateful). The Dutch case is given slightly more extensive treatment, as in this case I was able to have discussions also with Foreign Ministry staff beyond the evaluation unit itself.

3.4 In considering quality, I have had in mind not merely the quality of an evaluation study itself, as far as such a thing can be measured, but also its impact. A high quality study of an issue of little practical concern or relevance, or a high quality study that leads (for whatever reason) to no consequences whatsoever, is of little interest. On the other hand, the greater the potential for impact, the more essential the quality of the study itself becomes. Clearly a high profile study which reaches erroneous conclusions is the worst of all worlds.

3.2 Setting a favourable environment for high quality evaluations: how does DFID compare?

3.5 Based on my survey, DFID sets an environment that is broadly similar to those of the comparator agencies, but there are some elements of good practice that it might usefully emulate.

3.6 A clear and supportive overall **policy** towards evaluation is clearly one positive factor, which in DFID's case the Management Board have just put into place. The policy document itself is longer and more comprehensive than those of the comparators, among which the Irish and Norwegian examples seemed to combine direction and brevity particularly well.

3.7 On the much debated question of **independence**, the establishment of IACDI (an innovation being replicated in the Netherlands and being watched elsewhere) has given a more independent cast to the UK system, which previously would have rated towards the lower end of relative independence in this group. The head of Evaluation Department is still in some respects in a weaker formal position than some of his peers in the group (for example, his Dutch and Commission counterparts report to Minister/Commissioner level), but the UK is at the independent end of the spectrum in IACDI's having the final word on the evaluation programme. None of the comparator organisations has as yet taken the World Bank route of forbidding their evaluation heads to work again for the agency. It was interesting that the Sida Evaluation Department is shifting the focus quite sharply from accountability towards lesson-learning with a view to improving its relevance to the agency, in a context where SADEV, as an autonomous agency reporting to government, provides a new and independent evaluation voice.

3.8 Where DFID's evaluation department may possibly be in a relatively weaker position is less in formal positioning than in the DFID culture –apparent from Burt Perrin's work – of **persistent and detailed challenging of evaluation**

findings by operational units. It is particularly difficult to get a good comparative feel for this aspect across the institutions met, but while lively debate is clearly not confined to DFID, the heads of evaluation that I met did not highlight it as a major issue of concern.

3.9 Another structural issue is the link – if any - between evaluation and **internal audit**. In most of the agencies surveyed, the reporting lines of evaluation and audit came together only at Permanent Secretary or (as in DFID) at Director-General level. DCI had a model where the head of evaluation was also head of internal audit – a structure that apparently works well and provides access and insights. Sida had had such a model for over a decade, but had recently discontinued it. In the Netherlands, Evaluation (IOB), internal audit and Inspection make up the Foreign Ministry's so-called 'Cluster of Control', the three heads meeting regularly but each pursuing their own mandates. In the European Commission, the evaluation and internal audit functions had, by contrast, little connection. While there are logical reasons for each of these options (depending in part on the size of the agency), it did not seem that this issue was linked in any clear way to the quality of evaluations.

3.10 It was clear in my discussions that the increasing interest of **external auditors** in value for money issues was a more immediate issue, notably in relation to the programming of evaluations and such audits.

3.11 It is not easy to make fair comparisons of **financial and staffing resources** across different agencies, because of the lack of truly comparable data. Annex 11 attempts however to give some core numbers: evaluation managers (or in SADEV and IOB, evaluators or evaluation team leaders), operational budget for studies, number of studies issued in 2008, size of bilateral ODA (chosen because, with the exception of SADEV, very little evaluation resources go into assessing multilateral institutions, which of course have their own systems and are also examined through MOPAN and other frameworks). It then gives some 'key ratios' drawn from this data.

3.12 On this basis, the following conclusions are apparent, **though all need careful examination because of the problems of getting truly comparable data:**

1. DFID's budget for commissioning studies is above average for the group in relation to the size of the bilateral aid programme.

2. DFID's complement of senior evaluators is below average in relation to the size of the bilateral aid programme. These evaluators are responsible for managing an above-average number of studies per year.

3. DFID's evaluations are relatively numerous but also relatively inexpensive (the policy on country programme evaluations is one reason for this result).

4. DFID appears to lead joint studies more frequently than the average.

3.13 It is encouraging that the DAC Evaluation Network agreed in June 2009 to a study of the systems and resources of the evaluation operations of its members. This should provide a more comprehensive basis of comparison and, hopefully,

data whose comparability are better assured than I have been able to assemble in the time available for this study. It would be useful if IACDI could examine the results of this work at a forthcoming meeting.

3.14 Undoubtedly, the **experience and expertise** of evaluation managers is a key factor for quality evaluations. Here DFID is among the weaker performers in the group. There seem to me to be strong grounds for those who manage evaluations to comprise a mix of staff, some with a background in research and evaluation, some with operational experience, and for these staff to stay in post for a relatively long period, so that at any time the Department has a 'college' of experience and expertise that can be shared, and a strong learning environment that in turn facilitates the induction of new staff. But The NORAD model of half the staff recruited from outside with an evaluation specialism and the other half internal staff, or the Dutch model of one-third external specialists, one-third really long stay Ministry staff for whom evaluation is their career anchor, and one-third regular Ministry staff both seem worthy of attention.

3.15 A key factor which DFID needs to address seriously is that of **length of posting**. In the Netherlands, even for those who are not 'long-stay' staff, a four year posting is the norm. Managing evaluations is a skill where experience counts. It was telling to be informed by the head of IOB/Netherlands that he had refused – with difficulty but successfully – to accept sub-standard work from recognised consultants, and had felt able to do so because he had confidence in the judgements of his evaluation managers. He said that he doubted if DFID staff, with their more rapid turnover, would have sufficient confidence to expose DFID to the risk of judicial action in this way.

3.16 **Conclusion**: DFID is reasonably placed in respect of its operating environment to encourage good quality central evaluations (on decentralised evaluations and reviews, see below) but could do even better by bringing in more staff with a background in research and evaluation, and by keeping evaluation managers in post for a four-year norm. It should be a high priority to put in place a plan to achieve these two linked objectives. DFID could also discourage undue second-guessing of evaluation judgements (as opposed to factual inaccuracies) by the units being evaluated.

3.17 In terms of resourcing, based on a necessarily limited comparison with the other agencies, DFID could consider either expanding the number of its senior evaluators or cutting the number of evaluations, and perhaps increasing the resources per evaluation. (However, the findings of the Perrin study indicate the need to keep value for money in view.)

3.3 Deciding topics for evaluations likely to encourage relevance and buy-in: how does DFID compare?

3.18 DFID has a system for selecting what to evaluate that is paralleled quite closely in most of the comparator agencies. Typically, such agencies have some idea of what they would like to evaluate over a period (the period of the Financial Perspectives for the EC, for example), and then work out and present to senior management an annual programme that reflects broad consultation with agency staff. DFID's system is not dissimilar, though of course the final programme is

now decided by IACDI after the Management Board has commented on the proposals from Evaluation Department. IACDI has also now built in a useful new element of public consultation, an element of good practice not so evident elsewhere.

3.19 All these systems do however risk prioritising what interests a few key people in the decision-making chain over what actual demand may be. It is therefore interesting to reflect further on the Dutch system, which aligns itself more explicitly than the others which I reviewed with the objectives that drive the agency as a whole.

3.20 The Dutch objective-setting system is very similar to the Public Service Agreement concept used in the UK. For each budget process, activities are grouped under a series of 'Policy Articles' (8 in total for the Foreign Ministry last year), which are themselves then broken down into a few 'operational objectives' which effectively drive the business. For each of these, there has to be a policy review, typically on a 6-year cycle, which is to be supported by impact evaluations in the years preceding each policy review.

3.21 The Dutch approach means that by May each year, there is a draft rolling plan of evaluations for the following budget (calendar) year and two further years. The presentation also shows what is planned for the current year and what was done in the two preceding years, again categorised by operational objective. In this way, the evaluation programme supports rather directly the way that the business of the Ministry is run (IOB conducts evaluations on any Foreign Ministry programme, not just development cooperation). This should logically enhance the likelihood of evaluations being seen as relevant. While I understand that DFID's own rolling programme of evaluations in practice covers most PSA targets, I am not sure that it has this degree of integration into the business planning of the Department as a whole.

3.22 A second virtue of this regular and structured planning is that it enables the IOB to plan and put in place **impact evaluations** of key topics amenable to such approaches in a timely way. They have recently done this, for example, in education in both Uganda and Zambia with the governments of those countries – evaluations that with others to come will feed into a major policy review planned for 2011. This responds to the pressure that almost every agency is currently feeling to demonstrate results, in spite of which the other evaluation departments surveyed were commissioning virtually no impact evaluations, often on the grounds that the higher level issues that they needed to address could not be tackled in this way. The same is true of DFID, where the few impact evaluations available have been commissioned by country offices rather than by Evaluation Department (e.g. a £5 million grant to the Indian National Council for Education Research and Training NCERT and the London Institute of Education to study impact of a major co-financed programme in basic education). The Dutch approach – and IOB is looking to increase impact evaluations sharply from what is still a low level – shows that with good planning at least some aspects of the wider issues which central evaluation units are often called upon to examine are susceptible to more rigorous evaluation methods which will inform consideration of future policy.

This is not to deny that most impact evaluations could and should be commissioned, as in the India case, by developing country governments with support from country offices, or perhaps, as noted above, through research programmes.

3.23 It would of course be a mistake to make forward planning the enemy of responsiveness to changing circumstances, and the Dutch system, like the others examined, is adjusted on an annual basis to reflect all relevant factors. (For example, the Dutch Parliament has itself suggested issues for inclusion.) However, the scope for more explicitly linking DFID evaluations to PSA and other aspects of results-based management in DFID's own rolling forward plan may be worth considering, in particular to encourage a clear sense of ownership by the Management Board and by those responsible for target delivery.

3.24 As noted in the previous section, views differ among comparator organisations on the case for evaluating **country programmes**.

3.25 Views are closer on the issue of evaluating **multilateral aid**. Most comparator agencies draw the line at evaluating the effectiveness of their multi-bi (i.e. special purpose) contributions, as with a recent EC evaluation of funds transferred to IBRD and EIB. They reasonably regard the task of evaluating the overall performance of multilateral agencies as something to be assessed by a combination of the multilateral agencies' own evaluation systems and by multi-donor approaches such as MOPAN. SADEV stands out as an interesting exception, a position it is better able to take as a non-funding agency. Its critical report[8] on the quality of UNDP's in-country evaluations is a useful contribution to a more detailed understanding of what reliance may be placed on some agencies' own evaluation systems.

3.26 **Conclusion**: DFID's approach to selection of evaluation topics is in line with most comparators, though it is in the minority in its approach to country programme evaluations. It might however consider a closer link to the way that DFID objectives are set, supported by timely impact evaluation where relevant.

3.4 Involving partners and other donors in evaluations: how does DFID compare?

3.27 This question is of course closely linked to the last. A good deal of domestic discussion of evaluation in all countries still reflects a world where donors largely ran their own projects. In a world where a large proportion of aid funds are being committed for the joint financing of common objectives, and where host country ownership is rightly being recognised, this approach has to change, including in relation to evaluation. The Accra Agenda for Action commits participants to "better co-ordinate and link the various sources of information, including national statistical systems, budgeting, planning, monitoring and evaluation" (my underlining). The impact of joint studies, and

[8] SADEV Report 2008:6. Strengthening the Results-Orientation in Sweden's Engagement in Multilateral Development Cooperation. An evaluation of UNDP's country level evaluation activities.

studies with strong local ownership, can in principle be larger than that of studies by a single donor. But it will then be all the more important to safeguard their quality.

3.28 In practice, of course, systems are adjusting. **Joint evaluations**, a rarity before the Rwanda Emergency evaluation of 1996, now account for 50% of planned evaluations by NORAD and even more for Sida. Studies either commissioned by developing countries (as with the evaluation of the Poverty Eradication Action Plan in Uganda), led by developing country consultants (as with a NORAD study of NGO programmes in Northern Uganda), or conceived and executed as a genuinely joint effort (as with the Chinese/Dutch study of Dutch mixed credits to China discussed in the previous section) are no longer a total rarity: indeed their weight is certain to grow rapidly, despite the many practical difficulties.

3.29 DFID, with its large portfolio of joint finance, notably in the form of budget support, has played a strong role in promoting and managing joint evaluations, and is at the forefront of this trend.

3.30 While all the comparator agencies apart from the EC (whose financial arrangements make collaboration with other donors particularly difficult) and SADEV (though SADEV advise that they plan to do more in this direction) participate in many joint evaluations and expect to continue to do so, they all recognise the typically heavy transaction costs and the problems of over-complex 'committee design' and competing approaches to methodology and quality control. In this area as in others, therefore, a harmonisation effort is a high priority, the DAC EVALNET being the obvious venue for progress as among the bilaterals. A recent SADEV study suggested a number of areas where planning and management could be improved, which EVALNET has supported with the result that an interactive platform is going to be developed and tested by members. It is therefore important that IACDI and similar national institutions help to build international good practice rather than insist on separate national standards e.g. for quality control. (The same is true of other standard-setting bodies such as the Boards of the International Financial Institutions.)

3.31 As the aid effectiveness debate has taught us, harmonisation, however important, is less transformational than alignment. In the field of evaluation, the world is in fact very far from the AAA vision, quoted above, of '**country-led evaluation'**. Nevertheless, the direction of travel should be clear, towards a future where most developing countries will have some capacity to evaluate the effectiveness of key programmes (not just those financed by donors) and where donors will increasingly be able to use host-country led evaluation material to satisfy their need for knowledge of results and for lesson-learning, leaving their own evaluations to concentrate much more on their own effectiveness as partners in the enterprise of development.

3.32 From this perspective, the most significant development that I have come across in this study is the fact that a DFID country office is in discussion with a major host country about a substantial capacity-building contribution to enable that country's planning agency embark on an ambitious programme of impact

evaluations. This appears to be the result of specific interest by the relevant DFID office head, rather than of any centrally-mandated policy from DFID. Yet this is exactly the sort of initiative that the international donor community needs to be taking given the importance of high quality and effective public spending programmes in poor countries, not least those in receipt of budget support. Such initiatives need to be financed by country and regional programmes, using advice from evaluation units as necessary, not by the evaluation units themselves (UTV/Sida does have some modest earmarked funds for capacity development, but this was untypical.) **DFID and other donors should promote such approaches more actively.**

3.33 What evaluation units can do (and most of those contacted for this study are trying to do) is to do more to engage host governments in their evaluations and make more use of local consultants and universities to participate in and indeed in some cases to lead evaluations. Examples of good practice (which could be matched from DFID experience) included the Norwegian Embassy in Kathmandu encouraging NORAD to fund a Nepalese Government evaluation of a major education sector programme (an evaluation admittedly foreseen by the donors as part of their sector support), the Dutch evaluation agency involving the Uganda Ministry of Education closely in an impact evaluation of basic education, which led to the Ministry setting up its own evaluation unit to continue aspects of the work, the EC using a joint team of Tanzanian and European experts to conduct its country programme evaluation of Tanzania, and NORAD tendering an evaluation of CSO support in Northern Uganda locally, with the result that a local consulting company carried out the study. Together with broader capacity-building initiatives (such as the World Bank's support of regional centres of excellence), such activities will help to shift the still very donor-oriented evaluation culture in the direction that it needs to go. This is not to deny the often significant costs of greater reliance on local resources, nor the enhanced need for quality assurance that is likely to be needed. Alignment can only move at a certain pace.

3.34: **Conclusion**: DFID has an appropriate policy stance on joint evaluation, in line with most comparator agencies. Standard-setting bodies, including the DFID Management Board and IACDI, need to build harmonisation into their requirements. DFID should consider encouraging country offices to include support for evaluation systems (and for impact evaluation) in their dialogue with host countries on the effectiveness of public spending. DFID's Evaluation Department, like its comparators, should continue to encourage host-country led evaluation and use of local evaluation capacity (not just in a data collection role). It might experiment with local tendering of some evaluation work.

3.5 Managing evaluations with a view to a quality product: how does DFID compare?

3.35 Among the comparator organisations, Irish Aid, NORAD, UTV/Sida and EC all, like DFID, employ an 'evaluation management' model, whereby studies are carried out by consultants/institutions and managed by a member of the evaluation unit. SADEV uses a different model, similar to that of the World Bank, where SADEV evaluators and short term external experts form teams

(led by a SADEV staff member) to carry out evaluations. IOB/Netherlands has an intermediate model where each study team is led by a senior IOB 'inspector' but where much of the work is carried out by consultants under their direction: the inspector controls the whole process and takes a major role in the final report. As noted in Part I of this report, the sample of studies examined does not permit judgements to be made on whether any of these models is inherently superior to another. One agency argued that use of consultants was a first step to ensuring independence of judgement, but this must be questionable if consultants depend on agencies for future work.

3.36 All the agencies have, like DFID, which uses a version based on the DAC standards, some form of quality standards for the planning and carrying out of evaluations. NORAD and UTV/Sida apply the DAC standards; SADEV have recently on the basis of the DAC standards developed a quality assurance instrument for the agency's work, and IOB is working on new internal guidelines. The EC has the most developed and documented system – the four volume 'Evaluation Methodology for European Commission's External Assistance', which covers the full project cycle of evaluations with much specific detail.

3.37 The early stages of the project cycle are crucial for the quality of the final product. SADEV and UTV/Sida start each evaluation with a short memorandum/approach paper which is shared with the full team of evaluators, and discussed internally prior to formulation of terms of reference.

3.38 Establishment of some kind of 'reference group' is normal practice for all agencies, including DFID. The composition of such groups varies. All include relevant internal stakeholders, though NORAD consider that it is important that the stakeholders' groups are not steering groups, in order not to prejudice independence. Most also include stakeholders in the country concerned (or, for EC country studies, the Brussels-based Ambassador of the country in question). Some of these groups can be quite large (10-15 in the case of the EC). Irish Aid routinely includes an independent stakeholder from Ireland who is outside Irish Aid. NORAD distinguishes between stakeholders and Advisers, who are usually evaluation specialists and whose job is to advise NORAD on professional aspects of the evaluation (this is similar to DFID's current planning for quality assurance). SADEV, following academic practice, employs a mixture of peer review from within the agency and external referees.

3.39 Terms of reference are in most cases agreed by the reference group (and in Sida's case also by the Head of UTV). In SADEV, terms of reference have recently been introduced, and are discussed in an agency-wide seminar, and finally approved by the Director General. NORAD have experimented with more 'open' terms of reference, and had good experience with this approach in a particular study. My examination of specific studies showed some very good examples of terms of reference, such as the Danish country evaluation of Uganda, the Dutch evaluation of civil society advocacy and the Norwegian evaluation of HIV/AIDS, but some were less well focused.

3.40 Consultants are usually hired by comparator agencies, as they are by DFID, through international (sometimes national and occasionally local) bidding procedures, and many of the same consultants work for several of the agencies. Agency experience is that the quality of consultants is very variable, and most agencies find themselves balancing the need for the consultants to reach independent judgements with questions as to the evidence and the rigour in interpreting that evidence that underlies some of these judgements. Irish Aid are experimenting with an interesting model under which a team leader is contracted and one member of the evaluation unit and one member of the group being evaluated (but not the person responsible for the programme being evaluated) allocated to work as members of the team.

3.41 Important issues are usually determined at the inception phase. This, for example, is when in the EC system decisions are made on the evaluation questions (limited to a maximum of 10, and covering but not replicating the DAC criteria of relevance etc) and the 'judgement criteria' and indicators which will be used to answer them.

3.42 At the report stage, judgements on quality become crucial. The EC assesses quality against a grid of nine criteria ('relevant scope', 'sound analysis' etc) which are marked on a five point scale, and are publicly available. UTV/Sida employs a similar grid. As noted above, IOB/Netherlands have on occasion rejected consultants' work that they considered did not meet reasonable quality tests.

3.43 **Conclusion**: the similarity of approach among all the agencies which manage consultants is striking, though the EC process stands out for its detailed structuring. DFID is in line with the essentials of good practice, though it might usefully consider the EC approach of limiting the number of evaluation questions in order to avoid reports that are too long and discursive. IOB is perhaps exceptional in its readiness to take a tough line with under-performing consultants, but its concerns over variable quality were widely shared.

3.6 Disseminating findings in a way likely to secure impact: how does DFID compare?

3.44 All evaluation units aspire to disseminate findings in a way likely to maximise understanding and lesson-learning in the agency. A first step is, of course, to have a readable report, a matter covered, for example, in one of the criteria of the EC's quality assurance grid. Short summaries, as in DFID, are made by several agencies. The EC ensures that all reports go to internal stakeholders directly (perhaps 50 people). UTV/Sida sends a regular internal newsletter to all Sida staff, consistently with its focus on lesson-learning.

3.45 One incentive for reports to be taken seriously is to require those proposing new spending commitments to certify that they have consulted relevant evaluation material, as is done in Irish Aid, as well as in DFID. Another is to have some kind of senior level consideration of reports. The EC sends a 1-2 page summary of each report to the relevant Commissioner. In Irish Aid, the 'Coherence Group' of senior managers, chaired by the Director General, takes

all evaluation reports. In the Netherlands, IOB presents each report to the Director General and Directors at one of their weekly meetings. NORAD's Evaluation Department sends to the Secretary-General of the Ministry a memorandum a few weeks after delivery of each report which contains a summary of the report, a note on comments received, the Evaluation Department's own comments, and its recommendations on action to take.

3.46 It is common practice for seminars to be held, either internally or externally, though in some cases the evaluation unit looks to the evaluated unit to take a main role in this. In the case of SADEV, in addition to its own dissemination including to the Foreign Affairs Committee of Parliament, the Ministry of Foreign Affairs holds a workshop on each report. Seminars and workshops are often also held in country, as with the impact evaluations of IOB/Netherlands or the country programme evaluations of the EC.

3.47 There may of course be limits to how far agencies wish to promote evaluations that may be quite critical or which raise difficult issues. One agency mentioned a case where a high-profile and well attended seminar had been arranged by the evaluation unit in the face of resistance from the political level; another told me that its media department was not particularly keen to promote evaluation reports.

3.48 In all cases, as for DFID, all evaluation reports are published and are accessible on the very useful DAC-hosted Development Evaluation Resource Centre, DEReC. I conducted a brief review of the ease of access to evaluation reports from agency home pages. For DFID, in common with Irish Aid, SADEV and Sida, evaluation reports are found by going first to 'publications'. For the EC and the Netherlands, it is necessary to type in the word 'evaluation' in a search box (in the EC case leading to an excellent page for each study which includes, for example, the quality score for each of the nine dimensions used by the Commission). Only in the NORAD case is there an evaluation link on the home page itself, which indeed enables one to see the studies in progress or planned.

3.49 DFID has very recently, on IACDI's suggestion, instituted an annual report by the Head of Evaluation Department. Opinions vary on the utility of such reports. NORAD and UTV/Sida both produce them, and SADEV is planning to do the same. On the other hand the EC has discontinued its annual report and IOB/Netherlands does not produce one. One piece of good practice to my mind in the DFID version is the inclusion of selected studies by other evaluation agencies, such as the excellent report on "Managing Exit and Transformation" led by UTV/Sida with IOB, NORAD and DANIDA.

3.50 **Conclusion**: DFID is in line with the general practice. However one weakness compared to former DFID practice is that there is, since the demise of the former Projects and Evaluation Committee, no senior DFID committee which automatically discusses the findings and recommendations of evaluations. It would be desirable that all Country Programme Evaluations were discussed in the Country Programme Review Committee and all thematic evaluations in the Development Committee. By one means or another (perhaps an annual

meeting) the Management Board itself should have an opportunity to consider key lessons learned from evaluations. If evaluations are not worth official time at least at the level of the senior committees of the organisation, why commission them? On external promotion of reports, see below.

3.7 Following up recommendations systematically: how does DFID compare?

3.51 One has to start here with some realism: because an evaluation report makes a recommendation, it does not follow that management has to accept it. Rather, the test of a satisfactory follow-up system is that there is proper management consideration of all such recommendations, clarity over the decisions taken, and action to follow-up such decisions.

3.52 Based on these criteria, I found comparator performance to be mixed. In one case, the system for follow-up of recommendations was admitted to have broken down; in another a formal tracking system, based on that used for external audits, was about to be put in place; and in a third I was told that the seriousness with which recommendations were taken depended largely on reactions to the reports at the political level.

3.53 Nevertheless, the initial stage is usually clear. The highest profile is that of the Netherlands, where each evaluation is sent to Parliament by the Minister together with his policy response, though the process thereafter is less well documented, and an attempt to insist on a more structured system has run into concerns about undue bureaucracy (a wider current issue in the Netherlands' administration). In the Norwegian case, once the report has been submitted (see above) the department has 6 weeks to send a plan to the management of the Ministry. In Ireland, a written management response is needed before the file is closed, and is in fact usually made before publication. The EC system involves the drawing up of a 'fiche contradictoire', in which the relevant Service of the Commission has to state whether it accepts, rejects, or partially accepts each recommendation (See Annex 12 for the fiche for the EC country study of Tanzania).

3.54 In both the EC and Norwegian systems a report on action taken on the recommendations is required a year after the report. In the Norwegian case progress is reported in the head of evaluation's Annual Report, and in the EC case the 'fiche contradictoire' is readily available on the EC website.

3.55 In the other cases I was not able to find the management responses to individual studies via the website (except that in DFID's case they are integrated into the EVSUM). **It would be good practice in my view for all agencies to have an 'evaluation' button on their home page, leading directly to all published studies, (with EC-style quality scores), management responses to each study, and a list of studies in process or planned – the DAC Evaluation Network could perhaps set some general standards in this area.**

3.56 **Conclusion**: DFID's system for checking whether action is taken on recommendations has been improved with the introduction of checking by Internal Audit, and would now seem to be as robust as most of the comparators. However, the NORAD and EC practice of putting a summary of action taken into the public domain would be worth considering. DFID Evaluation Department should encourage DAC-wide standards for accessibility of evaluation reports and plans, together with management responses (this would also link to the International Aid Transparency Initiative). If, as suggested above, all evaluation reports are discussed by a senior committee, this would further ensure that senior level attention is paid to recommendations.

3.8 Outside Stakeholders' perceptions of evaluation products: how does DFID compare?

3.57 The conclusion of my discussions on this point with comparator agencies can be simply stated: in most cases, evaluations have little public resonance (much less so, for example, than external audit reports). The European Parliament, and some national Parliaments, have shown little interest. In some cases, including in Ireland, the press may carry short reports and some civil society organisations may take an interest in some reports.

3.58 The exceptions to this general statement usually arise because a report is seen as controversial. Thus in Norway, the Opposition made use of the rather discouraging analysis in Norway's country review of Zambia (see Part I above) to criticise government policy in Parliament.

3.59 In the Netherlands, Parliamentary interest in evaluations appears significant and more sustained than elsewhere. Only in the Netherlands are evaluation reports formally presented to Parliament, along with the Minister's 'policy response'. The Dutch Parliament has taken the initiative to propose that certain topics be evaluated, as with a major evaluation of all aspects of Dutch policy towards Africa. And Parliament has pressed successfully for the establishment of an Advisory Committee with a mandate not dissimilar to IACDI as a guarantor of the independence of IOB. This lively interest reflects the vigour of the debate in the Netherlands about the case for maintaining one of the largest aid programmes relative to GNI, which appears to be less of an issue in, for example, Norway and Sweden. But it may also reflect the willingness of IOB itself to take positions which counter agreed government policy (for example by roundly criticising the Ministry of Finance for collecting 100% of debt write-offs from the Foreign Ministry, but paying only 90-95% to the Dutch companies concerned, and thus pocketing the difference).

3.60 I am not aware of any detailed survey of the interest in the UK in DFID evaluation reports. It would however appear that the All-Party Group and the Select Committee between them have a somewhat greater interest in evaluation findings than in the Parliaments of comparator countries (and of Europe), other than the case of the Netherlands. My attempt to get feedback from UK CSOs suggested that they accord rather little interest to DFID's evaluation reports. I had one written and one oral response from four enquiries, neither of the responses indicating particular interest in DFID's evaluation

products to date. The most telling comment was, from one leading NGO, "Often we find the multilateral reports (UNDP, WB etc) more authoritative and better disseminated in terms of presswork – DFID doesn't seem to do much press work around its publications, so NGOs don't pick up on them". In the Management comments reproduced in Annex 4, it is worth noting the comment: "There is perhaps a question about whether the style of presentation has kept pace with the more accessible communications styles now used for some of our other products", but also the judgement: "Evaluations are not one of the primary ways in which DFID responds to the information needs of parliament, other parts of government or the public."

3.61 **Conclusion**: DFID evaluation reports probably rate average in terms of generating Parliamentary interest, and below average for civil society. It might be worth exploring further whether there should be a greater effort to interest the latter, particularly now that Evaluation Department has been given some additional resources for dissemination.

3.9 Handling evaluative material commissioned by other parts of the agency: how does DFID compare?

3.62 I found quite a lively debate on the issue of evaluations and reviews conducted by operating units of the agencies I visited. Three issues seem to be in play:

1) Is there enough, too much, or too little evaluative work going on in the agency outside the formal evaluation programme?

2) Does its quality and independence need to be enhanced?

3) What role if any should the evaluation unit play in relation to it?

3.63 The amount of evaluative work being commissioned by operational units is in some cases set by agency-wide rules. Thus in the EC external consultants are supposed to visit every live project costing more than Euros 1 million annually and at project completion and score it in the Commission's results-based management system. If this is being done conscientiously (and I was not able to check this, as the evaluation department no longer has any involvement in the process) there would be 1,400 such reviews a year by external consultants. Other agencies appeared to be rather less demanding, but I did not get comparable figures. (Indeed, in line with experience within DFID itself, most agencies do not appear to have a usable database of decentralised evaluative work.) DFID itself requires all activities over £1 million to be reviewed annually, but does not require these reviews to be carried out by external consultants, and does not characterise them as evaluations. On any basis, the volume of such work dwarfs anything attempted by any bilateral evaluation department. Most of it is presumably aimed at answering activity-specific questions, but these could range from 'Are project activities going to plan?' to, as we have seen, 'What impact is this programme having on basic education for x million children in India?'

3.64 Given this volume of 'reviews' and their potential range from annual monitoring of progress to full-blown impact evaluation, several agencies have been considering the quality of decentralised evaluative work. Three studies were brought to my attention.

3.65 I OB/Netherlands produced in 2004 a 'Quality Assessment of Evaluations Completed in 2002'. This followed a much earlier study of the same topic in 1995 which concluded that no systematic use was made of evaluation results in formulating policy, methodological requirements were not satisfied, there was too little involvement of partner organisations and local evaluators, and the role of monitoring was not clear. The 2004 evaluation looked at 83 reported 'evaluations', 27 by Ministry departments and 56 by Embassies. These were in the main interim or final evaluations of projects and programmes, forming part of the regular 2-4 year cycle characteristic of such activities. Around a quarter were performed by third parties or in partnership with third parties (often joint evaluations). The IOB team selected 25 representative evaluations for analysis of 16 indicators covering their validity, reliability and utility. It found that 6 scored well on 13-15 indicators, a middle group of 8 scored well on 9-12, while the remaining 11 scored fair or poor on more than half the indicators. Joint evaluations performed somewhat better than the average. The team found that many of the evaluations had the character of progress inspections, in some cases linked to the question of whether projects would be extended or not. They tended to look more at specific topics on which project officers needed information and were seldom so concerned with issues of efficiency and effectiveness. Terms of Reference came in for particular criticism.

3.66 UTV/Sida published a similar review in May 2008 with the title 'Are Sida Evaluations Good Enough?' Despite its title, this covered only decentralised evaluations. This study looked at 34 evaluation reports by decentralised units of Sida which, under the system then in force, were published in the Sida Evaluation series. The report basically concluded that the studies were not 'good enough' due to weaknesses in their terms of reference, inadequate consideration of some of the DAC criteria (notably efficiency), poor handling of attribution, weak choice of methodology, and insufficient focus on lessons learned. They scored rather better for relevance and for clear recommendations, though the evaluation team questioned the robustness of the evidence for these.

3.67 The EC Evaluation Unit also conducted a review in 2008. This covered 40 decentralised evaluations chosen in a randomised manner. It found that exactly half scored good or above and half scored poor or below. Applying the same methodology to central evaluations showed that 47 out of 48 scored good (or in three cases very good), showing a sharp division of quality between the two types of evaluation. The greatest deficiency against the evaluation department's quality grid was in what the Commission term 'defensible design', in other words the methodology chosen was not adequately explained or defended.

3.68 The three studies are reasonably consistent: decentralised 'evaluations' do not measure up to 'best practice' standards, and their quality is very variable. There is no reason to suppose that a study in DFID, or other aid agencies, would come up with radically different findings.

3.69 This does not of itself mean that it is a high priority to improve the quality of such 'evaluations' against the standards set for central evaluations, since it could be argued that they were reasonably 'fit for purpose' in terms of guiding management on how to take forward activities in a sensible manner and have different objectives from central evaluations. Indeed, it is interesting that:

1. Sida have reacted to their own report by deciding to cease to call decentralised studies 'evaluations' or to publish them in the Sida evaluation series (though they remain available on the website). They have therefore decided to term all such studies as 'Reviews', and not to subject them to the standards required for central evaluations.

2. The Netherlands Foreign Ministry considers that too many decentralised 'evaluations' are taking place. They see evaluation as a relatively expensive tool to be used where significant decisions are needed, and are concerned that a 'habit of evaluation' is leading to over-use of this tool.

3. The EC has made since 2000 a clear distinction between the evaluation of individual projects (seen as a function integral to the planning cycle and so a core task of operational units of EuropeAid) and the evaluation of the results of regional and sectoral policies, programmes and programming performance, which are seen as crucial to the success of external policies as a whole and where, to secure objectivity, independent evaluation under the auspices of the central evaluation unit is considered necessary.

3.70 Against this background, DFID's recent decision to advance pragmatically to the goal of higher coverage of the programme through decentralised evaluations appears judicious.

3.71 Finally, on the role of central evaluation units in helping to improve the quality of decentralised evaluative work, there is a range of practice: in the EC and IOB/Netherlands, advice is provided on a reactive basis, in the Dutch case as part of the schedule of one out of 14 senior evaluators. In both cases, the proportion of overall evaluation unit time the amount of support given would seem to be low. For Irish Aid and NORAD 20% of central unit time is said to be allocated to support decentralised evaluative work, and in Sida up to 30%.

3.72 **Conclusion**: DFID is poorly informed on what evaluative work is being carried on outside Evaluation Department. The decision to advance pragmatically towards more end-of-project evaluation looks manageable, and squares with concerns elsewhere not to reach for the tool of full 'evaluation' too readily. Nevertheless, the studies by IOB/ Netherlands, UTV/SIDA and EC show that sustaining quality of decentralised evaluations is not easy, and it will be necessary to allocate significant resources if the new system is to achieve its potential. It will also be necessary to consider how the expanded programme will meet the requirements of harmonisation and alignment indicated above.

ANNEX 1: ASSESSMENTS OF COMPARATOR EVALUATIONS CONSIDERED IN THE REVIEW

Evaluation	Assessment category	Excellent	Good	Acceptable	Borderline	Unacceptable
Evaluation of the European Commission's Support to the United Republic of Tanzania	Overall assessment		x			
	Planning and conception		x			
	Evaluation Design		x			
	Implementation			x		
	Analysis		x			
	Reporting		x			
	Utility and use		x			
	Overarching considerations		x			
Evaluation of the Irish Aid Country Strategy for Uganda	Overall assessment			x		
	Planning and conception			x		
	Evaluation design		x			
	Implementation			x		
	Analysis			x		
	Reporting		x			
	Utility and use		x			
	Overarching considerations			x		
Evaluation of Norwegian Development Support to Zambia 1991-2007: NORAD Evaluation Department	Overall assessment			x		
	Planning and conception		x			
	Evaluation design		x			
	Implementation			x		

Evaluation	Assessment category	Excellent	Good	Acceptable	Borderline	Unacceptable
Evaluation of Norwegian Development Support to Zambia 1991-2007: NORAD Evaluation Department continued ..	Analysis			x		
	Reporting		x			
	Utility and use			x		
	Overarching considerations[9]					
Danish Assistance to Uganda, 1987-2005	Overall assessment		x			
	Planning and conception	x				
	Evaluation design		x			
	Implementation[10]					
	Analysis		x			
	Reporting		x			
	Utility and use[11]					
	Overarching considerations			x		
Joint Evaluation of the ORET/MILIEV Programme in China: IOB Netherlands	Overall assessment			x		
	Planning and conception		x			
	Evaluation design			x		
	Implementation		x			
	Analysis			x		
	Reporting			x		
	Utility and use		x			
	Overarching considerations			x		

[9] Insufficient information to rank this criterion.
[10] Insufficient information to rank this criterion.
[11] Insufficient information to rank this criterion.

Evaluation	Assessment category	Excellent	Good	Acceptable	Borderline	Unacceptable
Evaluation of Swedfund: SADEV	Overall assessment				x	
	Planning and conception				x	
	Evaluation Design			x		
	Implementation			x		
	Analysis				x	
	Reporting				x	
	Utility and use				x	
	Overarching considerations				x	
Promoting Women's Human Rights and Enhancing Gender Equality in Kenya: SIDA African Department	Overall assessment			x		
	Planning and conception		x			
	Evaluation design			x		
	Implementation		x			
	Analysis			x		
	Reporting			x		
	Utility and use		x			
	Overarching considerations			x		
Evaluation of the 'Strategy of Women and Gender Equality in Development Cooperation (1997-2005)': NORAD	Overall assessment		x			
	Planning and conception		x			
	Evaluation design			x		
	Implementation			x		

Evaluation	Assessment category	Excellent	Good	Acceptable	Borderline	Unacceptable
Evaluation of the 'Strategy of Women and Gender Equality in Development Cooperation (1997-2005)': NORAD continued	Analysis		x			
	Reporting		x			
	Utility and use		x			
	Overarching considerations		x			
'Turning Policy into Practice : SIDA's Impmentation of the Swedish HIV/ AIDS Strategy'	Overall assessment		x			
	Planning and conception		x			
	Evaluation design		x			
	Implementation[12]					
	Analysis		x			
	Reporting		x			
	Utility and use		x			
	Overarching considerations			x		
Evaluation of Norway's HIV/AIDS Responses : NORAD's Evaluation Department	Overall assessment			x		
	Planning and conception			x		
	Evaluation design		x			
	Implementation			x		
	Analysis			x		
	Reporting			x		
	Utility and use			x		
	Overarching considerations			x		

[12] Insufficient information to rank this criterion

ANNEX 2: AGGREGATE RATINGS - NON-DFID EVALUATIONS

Assessment category	Excellent	Good	Acceptable	Borderline	Unacceptable
Overall assessment of the evaluation	**0**	**5**	**6**	**1**	**0**
Planning and conception	2	7	2	1	0
Evaluation design	0	7	5	0	0
Implementation[13]	0	3	7	-	0
Analysis	0	6	5	1	0
Reporting	0	6	5	1	0
Utility and use[14]	0	7	3	1	0
Overarching considerations[15]	0	5	5	1	0

[13] Two studies not scored (insufficient evidence)
[14] One study not scored (insufficient evidence)
[15] One study not scored (insufficient evidence)

ANNEX 3: NORWAY - ZAMBIA EVALUATION EXECUTIVE SUMMARY

Background

This is an evaluation of Norwegian development cooperation with Zambia between 1991 and 2005. The context for the evaluation is a perception of the ineffectiveness of aid to Zambia in general, and Norwegian aid in particular, in addressing the continuing, and at times deteriorating, levels of poverty in the country. The evaluation team was asked to explore power relations within Zambia and between donors and Zambians with a view to answering the following questions:

- How do power relations within Zambia and amongst the donor community affect the scope for achieving pro-poor outcomes through development cooperation?
- How do power structures at different levels of Zambian society influence the potential of aid to reduce poverty?
- How do domestic and international power relations in Zambia affect the outputs of
- Norwegian-financed pro-poor interventions?

The evaluation, which was carried out between May and October 2007, had two phases; a study of power structures and development cooperation as related to Zambia and three case studies of Norwegian support to wildlife management, transparency in financial management and agriculture in Northern Province.

Approach

The evaluation tries to bring together the DAC evaluation criteria of relevance, effectiveness, impact and sustainability with an analysis of formal and informal power and political structures. The criteria is used to examine why the objectives of aid policies and development cooperation may have been undermined by formal politics, informal networks of actors and personal interests, leading to lack of progress in achieving sustainable development and poverty reduction.

To do this, the team examined different forms of documentation: policy statements, project documentation, existing evaluations of Norwegian and other donor programmes, studies commissioned by donors and academic studies. In addition, interviews were held with a variety of key stakeholders, including those who had been active in Zambia in the 1990s. Focus group discussions were carried out as part of the case studies.

Power Structures

Ideally aid is used in the ways outlined in locally owned development-policy documents and strategic plans; it augments domestic funds and supports initiatives that strengthen formal institutions and empower local populations so that government and citizens work together within a comprehensive, consistent, and well-articulated national policy and framework to generate economic growth and to distribute gains equitably to alleviate poverty. This model can fail for a large number of reasons and in various ways. The aim of this analysis of political power and structures is limited though, to exploring only one of these – the 'logic' of Zambian politics, the way it has

changed over the course of three regimes, and how it has impacted aid effectiveness. Here we focus on the Zambian elite because they have been relatively unrestrained by citizens until recent years.

In explaining why the political system may operate in ways detrimental to development and the poor we have taken a cue from political scientists writing about Zambia who have labelled the nation 'a neopatrimonial' state. We examine the evidence that Zambia does fit this model, which indicates a tendency towards the centralisation of power in the hands of a president and a small group of elite, who as 'big men' sit atop their own clientelist networks and are dependent on him and his access to state resources. They are rarely held accountable by a (relatively) weak civil society or by formally constituted but under-funded representative and watch-dog institutions that are prone to 'capture' by those they are supposed to be holding accountable. Into this accountability vacuum donors have stepped, trying to hold governments to account through aid conditionalities and by writing policy themselves. We conclude that, overall, this provides a good explanation of Zambian power relations during the period of the evaluation. However, there have been changes over the latter part of the period which point to a fragile move towards a more rational-technical state with a stronger developmental agenda.

The choice of development models used by Kaunda, Chiluba and Mwanawasa has been influenced by donors, contemporary economic theories and African-nationalist agendas, and by each man's desire to stay in office, leading to their use of public resources and the advantages of incumbency. In due course all three adopted economic liberalisation and structural reform as national policy, but each regime demonstrated a reluctance to implement some of their key provisions. In general this is because *implementing* – as opposed to *writing* – policies can actually undermine the imperatives or 'logic' of clientelist politics.

This is not to say that the failure to reduce poverty or consolidate democracy can be blamed solely on Zambian politics, for international economic factors, vacillating donor demands, and regional politics also have an impact. But domestic political systems and structures are key in affecting the way national resources are allocated, including how meritorious and incorruptible the civil service is, whether capacity building is prioritised and how strong the media and watch-dog agencies are. In the absence of a strong political-party system, an independent parliament and watchdog organisations, NGOs, community-based and faith-based organisations have increasingly formed 'the only serious issue-oriented political opposition in the country'. That said, the bulk of the rural population -- more isolated, impoverished, and uneducated – appears to have remained marginalised, under-served by the state, and voiceless. A strong and vibrant media has also developed over the years, and has at times held the government to account.

During the last forty years the overall structure and functioning of Zambia's political system has changed slowly, and this has affected the course of development and the ways in which aid could be used. Donors are beginning to understand that because transformation has to be generated domestically by deeply rooted forces their impact can be marginal at best. It is important that they put aside a purely normative approach and include political-economy analysis to support their interventions.

International Donor Assistance

Throughout the period of the evaluation, Zambia has been one of the major aid recipients in Africa. For the earlier period, in the 1990s, multilateral donors, and in particular the IFIs, were the most important donors. However, since 2001 bilateral flows have more than doubled, while multilateral flows have fallen to 11% of the total by 2005.

In the 1990s, development cooperation programmes were negotiated on a one-to-one basis with government, and there was little formal attempt to coordinate assistance amongst donors except through the Consultative Group meetings, which started in the mid-1980s and focused on macroeconomic and structural policy reforms. Some donors did act together on certain issues, such as the constitutional changes banning Kenneth Kaunda from standing in the 1996 elections. Governance became a much more prominent element of donor dialogue from the late 1990s onwards.

The dynamic of donor engagement with the Zambian government has changed since 2001. The CG process had a hiatus over the election period, and in September 2002, the Nordic Plus countries met in Helsinki and agreed to push forward a joint harmonisation process in Zambia. This was taken forward by the embassies in Lusaka, and the "Harmonisation in Practice" initiative, HIP, was started, which has, in turn, led to the process for the Joint Assistance Strategy for Zambia. Over time, this alignment process has incorporated more agencies, including the multilateral agencies, the WB and the UN system, and is now known as WHIP, the Wider Harmonisation in Practice. It remains to be seen whether, as some stakeholders fear, harmonisation and the need for consensus will constrain those donors prepared to take a stronger stance on issues of governance.

Norwegian Development Cooperation

Zambia has been one of Norway's partner countries over the evaluation period. Most assistance has been given on a government to government basis, though this has fallen from almost 95% at the beginning of the period to a low of 69% in 2001 with concerns over the election, rising back to just under 80% in 2005. Assistance to local NGOs rose to 20% of the total in 2001, but now Norwegian NGOs receive over 15% of assistance, though this is often disbursed to local NGOs. Since 2000 education, agriculture, environment/wildlife, health - basically HIV/Aids related,, financial management and good governance have been the main sectors receiving support.

Norwegian development policy has evolved since 1992, when the concept of recipient responsibility was introduced. There has been more emphasis on working with the private sector and civil society, on improved donor coordination and on support to MDG achievement. Currently Norway is gradually reducing the number of sectors with which it is involved to two or three per country, plus budget support.

Norway's engagement with Zambia has been a combination of policy dialogue, and project and programme support using a number of different partners and modalities.

An analysis of the policy dialogue from the 1990s, looking at minutes of consultation meetings between the two governments, plus CG meetings, shows that this was focused very much on macroeconomic issues. Norway was not unusual in its concentration on issues of liberalisation and privatisation. There was little room for discussion around the principles of economic policy, only on the implementation. Discourse analysis reveals a recurrent emphasis on the believed technical correctness of such policies and, insofar as problems are identified, these are located within the realm of tardy implementation. Overall there was a tendency to depoliticise economic issues, and address political objectives, such as the fight against corruption, separately, leading to potential incoherence between objectives.

Modalities of assistance changed over the period, with a move away from projects towards greater support to government sector programmes. Examination of support to NGOs and UN initiatives show that there have been considerable problems of sustainability as projects have faced difficulty in integration within wider government service delivery systems.

The case studies show that, over the period of the evaluation, there has been pressure on projects to adopt a more centralised management model. An analysis of the eleven projects included in the evaluation leads to the possibility that harmonisation is leading to a concentration on the two big players, government and donors, at the expense of CSOs and NGOs. The place for third parties at the development cooperation table is becoming smaller, and more contested. The effect on the space for pro-poor development cooperation is difficult to assess, but may end up with more importance on policy dialogue.

Development Assistance, Poverty and Power Structures

The analysis of the power relations indicates that there have been no obvious champions of the poor in national government, at least since the Kaunda period. Donors have had two major formal channels with government for influencing this: through policy dialogue and through direct funding of projects with poverty alleviation as a major objective.

Policy dialogue was focused on the macro-economy for much of the 1990s, and very much at a technical level, though with concerns arising towards the end of the period about governance. The development of the PRSP as a policy instrument by the IFIs created an opportunity for civil society to engage more directly with government and also with donors, but the line ministries did not, in general engage with or 'own' the process. When Zambia moved to the FNDP, as opposed to the PRSP as a planning framework, the involvement of the line ministries increased considerably but the process caused frustration at the provincial and district level.

The evaluation team found that Norway, in line with the approach taken by other donors, has tended to focus on technical issues in its formal dialogue with government. It has been cautious in addressing issues which could be seen as overtly political, where the legitimacy of its engagement could be challenged. Rather it has focused more on building capacity for Zambian organisations, both within government and with civil society partners, in the expectation, or perhaps the hope, that this will both build challenges within the system to abuse and will develop champions for the poor.

In this respect Norway has taken a similar approach to other donors. Insofar as internationaldonor personnel analyze the political element of economic reform, they tend to do so from the perspective of 'strengthening the domestic constituency for reform', promoting 'country ownership' of reform programmes, and creating the conditions through which governments can 'build consensus' for reform. The actual content of reform – based as it is on assumptions of a single, 'correct' approach – is often assumed to be beyond argument, and the task of politics is simply to persuade people of the merits of implementing reform. Poverty reduction becomes depoliticised, and there is some evidence that poverty reduction lacked both national and international political prioritization in Zambia.

Norway as a Donor

The perception of Norway as a donor is often difficult to distinguish from a perception of the Scandinavian donors as a whole, who are seen as supportive, in part because of their history of support to Zambia as a frontline state in the past. There is little knowledge in Zambia of the precise nature of Norwegian support currently.

Norway is seen as a flexible and generous donor, though there was a perception that Norway was not as rigorous in its approach as other donors, focusing more on process and less on results and impacts. Perceptions of how well Norway has encouraged greater ownership of projects are mixed. Views have been expressed that projects started in the 1980s and early 1990s were very much seen to be designed by expatriate organisations and consultants with limited input from Zambians themselves. This perception has changed as Norway's projects have both become more aligned to national approaches and support has moved more towards either sector approaches, or larger multi-donor projects.

The issue of ownership is more difficult to disentangle. This is largely because for many of the larger projects the relevant question is not is there national ownership, but *whose* ownership within Zambia. For some projects, this is an issue of layers within the hierarchy – there may be ownership at national level, but much less at the level of implementation, and even less at the level of the ultimate intended beneficiaries. In the area of support to improved financial governance, it is more difficult to pin down ownership in these terms. Rather the various projects are seen as supporting reformers within government, who have a vision of a more self-reliant Zambia, free from corruption, and are therefore promoting a joint agenda with donors and these reforming elements.

Norwegian support appears to score well on relevance. Analysis of policy documents, including MoUs, indicates that Norwegian development support to Zambia has become more aligned to Zambian policies and strategies over the years. There is evidence of greater consultation within Zambia on areas of mutual concern, and over the years there has been more assistance given in the form of sector and budget support. The recent process of alignment has taken this a step further, with Norway delegating support in some areas through other development partners.

The evidence on effectiveness is more mixed. A relatively poor level of monitoring during implementation has reduced project effectiveness, and Norwegian support has been most effective in delivering intermediate outcomes such as capacity development and institutional strengthening.

The evidence from the case studies on impact and sustainability show that much depends on the effectiveness and robustness of the government organisations involved in the projects.

Support to the Auditor General's Office appears to have had positive impact, but it is too early to comment on sustainability.

Conclusions

Norwegian aid to Zambia shows a similar pattern to that of other donors, in that it went through the same processes of emphasis on firstly macroeconomic stability and structural adjustment in the 1990s, followed by a focus on governance, but in an apolitical manner, emphasising reform of state institutions. For the most part bilateral donors have moved in the same direction, though at times at different speeds. Recent initiatives on harmonisation are reinforcing this tendency and also now bringing in the multilateral donors and IFIs into one process with the bilaterals.

There has been success on the macroeconomic front, in part due to external factors. The focus of dialogue has moved to governance issues, again with some success. There appears to be some Zambian effective demand for improved governance, not just in civil society organisations but also at technical and political levels within government. Such reform as has taken place has been within the constraints arising from the strongly centralised and personalised nature of power in Zambia, and as such they are rather fragile in nature.

However they are an important step in the direction of developing a technical-rational modern state, rather than a patrimonial one.

To support the development of a modern democratic state in Zambia, Norway could take the following actions:

- Improving its own understanding of the political context of Zambia.
- Supporting the further development of formal checks and balances in political life.
- Funding capacity building in support of decentralisation.
- Including civil society in the harmonisation process.
- Increasing support to local NGOs rather than through Norwegian organisations.
- Encouraging intellectual and public debate on political issues and on poverty reduction.

ANNEX 4: DAC WORKING TOOL FOR PEER REVIEWS AND ASSESSMENTS: EVALUATION SYSTEMS AND USE

E = DFID Evaluation Department response
M = DFID Management response – DG Corporate Performance
OS = Responses by Other Stakeholders (civil society organisations)

1. Evaluation Policy: Role, Responsibility and Objectives of the Evaluation Unit

1.1 Does DFID have an evaluation policy? (E)

Yes, recently approved in March 2009.

1.2 Does the policy describe the role, governance structure and position of the evaluation unit within the institutional aid structure? (E)

Yes. It describes the role and mandate of evaluation. It explains how evaluation fits with the rest of DFID's accountability, results and performance management framework and external scrutiny arrangements. It sets out how Head of Evaluation interacts with key governance arrangements in DFID.

1.3 Does the evaluation function provide a useful coverage of the whole development cooperation program? (M)

Evaluation Department carries out a comprehensive set of country programme evaluations (on a rolling programme in which each country is covered every few years). It has evaluated implementation of most of our major policies, including our approach to the use of particular aid instruments, such as budget support and technical cooperation. It has also evaluated at least one major DFID research programme.

Key areas of our work (fragile states, budget support, governance, conflict, HIV and gender) have been well covered in thematic and country studies. Climate change has been a bit of a gap but is now on the work programme.

Evaluation Department has not systematically evaluated the effectiveness of money provided through multilateral agencies or civil society bodies, where we generally rely on the partner's evaluation system.

For the past few years, the central evaluation function has not had responsibility for monitoring and evaluation of individual projects. We have recently recognised the need to strengthen evaluation of individual projects, and our new policy provides a larger role for the evaluation function in ensuring country offices commission independent evaluations of a larger range of projects. Our Evaluation Department has also directly supported impact evaluations of a small number of project interventions.

1.4 According to the policy, how does evaluation contribute to institutional learning and accountability? (E, M)

Institutional learning and accountability form the two-fold core of DFID's Evaluation Mandate, set out in the Policy (Box 3). Evaluation is described as an "essential element of DFID's accountability and performance management framework" and as a "key instrument both to inform decision-makers…delivering credible and impartial findings and recommendations which are reflected in DFID's policy and programming…and to hold DFID to account for its choices and actions."

Our new policy cites learning and accountability as the twin goals of evaluation.

Evaluation contributes to learning through a range of mechanisms, including:

- Direct feed through from evaluation of policies into future policy development (e.g. evaluations of HIV/AIDS, gender, budget support), and from evaluation of country programmes into future country planning (e.g. Afghanistan, Yemen);
- Publication and dissemination of evaluation findings through DFID;
- Transmission of knowledge through advisory cadres;
- Role of Head of Evaluation on key corporate committees (development committee, investment committee etc).

Evaluation also contributes to accountability through a range of mechanisms including:

- Publication of evaluations;
- Summary of evaluation findings presented to management board;
- Tracking of follow-up to evaluation recommendations.

1.5 How is the relationship between evaluation and audit conceptualized within the agency? (E, M)

DFID's monitoring, evaluation and audit functions work together to assess results. Project results are monitored 'in the line'. Systems for results management in DFID are being strengthened through the Results Action Plan and Making it Happen initiatives, supported by the Value for Money Department, scrutinised inter alia by the Audit Committee and Investment Committee. Audit assesses systems, procedures and risk management. Evaluation uses and challenges monitoring information to independently assess relevance, efficiency, effectiveness, impact and sustainability of DFID's policies and interventions.

The functions are seen as distinct and separate – with evaluation focused on the development impact of our policies and programmes; and audit focused on adequacy of systems and procedures, quality of compliance and management of risk. The two functions are currently co-operating on an evaluation of our anti-corruption work.

1.6 In countries with two or more aid agencies, how are the roles of the respective evaluation units defined and coordinated? (E, M: in the UK case, this would be about EVD and IACDI)

DFID is the UK HMG's main aid agency. However HMG's international development objectives are increasingly achieved through collaborative working with other government departments including e.g. FCO, BERR, DEFRA. The FCO is just establishing its own evaluation department with whom EVD is in contact; the

Government Economic Service has recently established a Whitehall Evaluation Focus Group, in which EVD participates; under the new Evaluation policy more formal protocols for ensuring OGD engagement in DFID evaluations will be developed. These will formalise current practice for cross-Whitehall engagement particularly on evaluation of development issues with joint PSA objectives where policy coherence is essential e.g. on engagement in fragile states, international trade.

The UK has only one aid agency, so the question about coordination between different evaluation units does not apply.

In 2007, the Secretary of State for International Development created an Independent Advisory Committee on Development Impact (IACDI) to strengthen the independence of the evaluation function and assure the quality of its work. The Committee reports to the Secretary of State, and plays a strong role in the setting of Evaluation Department's policies, systems and work programme.

1.7 Is the evaluation policy adequately known and implemented within the aid agency? (E, M)

The drafts of the policy have been widely discussed through EVD's internal consultation process over the last year, including a cross-departmental Policy Steering Group chaired by the DG Corporate Performance, and by DFID's Development Committee chaired by DG Policy, and by DFID's Management Board. The Secretary of State has recently approved the evaluation policy, and we will be internally and externally launching it through various channels e.g. press release, announcements on internal and external website, DFID's Aid Effectiveness Newsletter, and specialist advisory networks. One of the pillars of the Policy is to strengthen DFID's culture of learning and evaluation and starting 09/10 financial year, EVD will be investing strongly in evaluation capacity development across the organisation, in collaboration with the Results Action Plan, Effective Programmes Team and Making it Happen initiative. IACDI have been closely engaged with the development of the policy and have helped raise its profile significantly with DFID's senior management and ministers.

My sense is that our basic approach to evaluation is well known and understood within DFID, and has gained increased profile in DFID through the creation of IACDI, and the prominence of evaluation in our Results Action Plan. A survey conducted by the NAO for their recent study of monitoring and evaluation in DFID found that 79% of respondents were familiar with the role and mandate of EVD. There has been relatively wide consultation about our new evaluation policy, but the details of it will not yet be widely known and a communications effort on this is planned.

2. Impartiality, Transparency and Independence

2.1 To what extent are the evaluation unit and the evaluation process independent from line management? (E, M)

To ensure the evaluation function's independence is protected, it is overseen by IACDI and the Evaluation Dept reports directly to the DG Corporate Performance as a member of DFID's Management Board - a relatively recent change from its former reporting line through the Director of Finance and Corporate Performance.
Evaluations commissioned by EVD are managed independently of operational and line management. Operating departments and staff have the right to comment on factual accuracy and balance of draft reports, and final judgment is made by Head of Evaluation.

- Evaluation Department reports to Director-General Corporate Performance (i.e. straight into top management group, outside any Divisional structure);
- The current Head of EVD was appointed directly into Evaluation Department from outside the organisation;
- IACDI was created to buttress independence of the function;
- Evaluations are generally conducted by independent consultants;
- However, there is strong reliance on internal staff for evidence; and consultation with staff over factual accuracy and balance of reports. Evaluation Department is keen to engage with staff in discussion and dissemination of findings, to ensure that lessons are learned and recommendations implemented.

2.2 What are the formal and actual drivers ensuring / constraining the evaluation unit's independence? (E, M)

IACDI report to the Secretary of State; Head of Evaluation reports to the DG Corporate Performance. IACDI has made a significant and real difference to the positioning, profile and independence of EVD in its first year's operation (2008). IACDI made a number of recommendations to the SoS to strengthen EVD's independence, half of which are already implemented and several more are in hand or under active consideration. The UK's 2006 International Development (Reporting and Transparency) Act sets the legal basis for DFID's accountability, and the International Development Committee (IDC) regularly debate DFID's evidence and impact, meeting annually with IACDI. The NAO has recently completed a management review of DFID's monitoring and evaluation systems, in collaboration with IACDI.

Less formally, increased public scrutiny of the UK's aid budget and greater international commitment to the quality and impact of aid has driven DFID's recognition of the need for independent evaluation. The new Evaluation Policy will help clarify to all staff the purpose and mandate for evaluation in DFID, and clarify staff roles to ensure appropriate engagement in studies.

As set out in answer to 2.1 above.

IACDI is a strong actual driver for independence. The professional integrity of Evaluation Department staff and consultants is also an important factor. The Head of Evaluation Department regards it as a core part of his role to resist pressure to tone down messages, and to protect EVD staff and evaluators from undue pressure.

A substantial proportion of our evaluations are joint evaluations with other donors. This provides an additional protection of independence from DFID, because there is very little scope for a single partner to influence the findings.

As members of DFID, Evaluation Department staff are arguably heavily influenced by DFID's overall world view; and are subject to some implicit (or occasionally explicit) pressure to give DFID the benefit of the doubt on more critical conclusions. They also rely heavily on DFID staff for evidence and comment – this partly reflects the constraints on time and other resources for collecting independent evidence.

2.3 What is the evaluation unit's experience in exposing success and failures of aid programs and their implementation? (E)

- The Head of EVD writes an Annual Report considered by IACDI and Management Board which highlights key evaluation findings and makes recommendations;
- EVD and Internal Audit Dept track management responses to evaluation recommendations and report annually on these to IACDI;
- Country Programme Evaluations feed directly into country programme business plans through the Country Planning Review Committee. The head of EVD sits on this high level committee and on the Investment Committee;
- Thematic evaluations have consistently informed future policy and strategy e.g. the major joint evaluation of budget support led by DFID/EVD is now the standard reference on this issue in DFID and internationally. DFID updated its Guidance on Budget Support and Aid Instruments based on this evaluation's findings and recommendations. More broadly it has influenced policy thinking in UK through IDC's deliberations and the NAO report on budget support. The HIV evaluation has been used to update the UK strategy on HIV & AIDS, and the gender evaluation has contributed significantly to DFID's Gender Action Plan in 2007;
- All EVD commissioned studies are published. As noted earlier the Head of Evaluation has the mandate to decide on final wording after considering staff views on factual accuracy and balance of draft reports. Sometimes this process can be prolonged, lengthening the period between field investigation and final publication and thus reducing somewhat the timeliness and relevance of reports. However, often the evaluation process itself, through appropriate steering group engagement, has triggered action on the recommendations.

2.4 Is the evaluation process transparent enough to ensure its credibility and legitimacy? (E, M, OS)

All EVD studies are published; both internal and external stakeholders are involved in steering and reference groups. The Policy and Workplan has been subject to full public policy consultation process, and is scrutinised by IACDI.

Yes. All evaluations are published. The establishment of clear evaluation models for each study, and the use of independent consultants to conduct evaluations provides an additional assurance of transparency in the process. We do not have systematic

evidence on perceptions of users about evaluations – but my guess is that readers will generally regard evaluation conclusions as credible and well-evidenced, but may wonder whether some evaluations pull their punches on more critical conclusions.

The fact that the UK chairs the DAC Evaluation Network, leads the NONIE impact evaluation network of networks, and has led a range of multi-donor evaluations suggests that our evaluation function is regarded as a serious and credible one by peer agencies. The OECD/DAC peer review of 2006/7 referred positively to the rejuvenation of DFID's evaluation function.

Civil Society consultees did not comment on this aspect.

2.5 Are evaluation findings consistently made public? (E, M)

Yes, all EVD commissioned studies are published.

Yes.

2.6 How is the balance between independence and the need for interaction with line management dealt with by the system? (E, M)

The decision to set up IACDI while retaining an internal evaluation function was precisely aimed to achieve the appropriate balance.

Interaction with line management in DFID is very strong – with clear consultation and engagement at all stages in the evaluation cycle (planning, gathering of evidence, reaching of conclusions, follow-through into new policy). Evaluation staff are very much seen as part of the wider DFID team, and represented on all DFID's key committees.

Independence is protected by the mechanisms set out in 2.1 above. The question posed for DFID by IACDI and others is whether the balance in DFID is still too much in favour of interaction as opposed to independent.

2.7 Are the evaluation process and reports perceived as impartial by non-evaluation actors within and outside the agency? (M, OS)

A survey in the recent NAO review of DFID's monitoring and evaluation found that 25% of staff and 23% of other stakeholders rated EVD as entirely or very independent.

But my impression is that within DFID, Evaluation Department is generally seen as a highly professional function taking a critical approach to our work. I would hope the same applied to external actors – although see 2.4 above.

Again, civil society consultees did not raise issues on this point.

3. Resources and Staff

3.1 Is evaluation supported by appropriate financial and staff resources?(E)

The DAC Peer Review of UK Aid 2006 concluded "The Evaluation Department (EVD) is undergoing a process of rejuvenation which includes additional administrative and programme resources. The Department now contains 20 staff and a richer skill mix, and has more direct linkages to senior management and DFID decision making. …". IACDI is engaged in discussion with Senior Management and Ministers with the aim of increasing resources for evaluation, to ensure the new Policy is fully delivered. Staffing and skills are a serious issue and is being prioritised this year with the identification of evaluation as a DFID specialism, the development of an evaluation competency framework, a training plan and a recruitment drive.

3.2 Does the evaluation unit have a dedicated budget? Is it annual or multiyear? (E)

Yes, it has a dedicated budget framework set indicatively for 3 years and reviewed annually.

3.3 Does the budget cover activities aimed at promoting feedback and use of evaluation and management of evaluation knowledge? (E)

Yes.

3.4 Does staff have specific expertise in evaluation, and if not, are training programs available? (E)

Yes to both and as part of implementing the new Evaluation Policy, EVD is set to raise specialist skills in EVD and across DFID.

3.5 Is there a policy on recruiting consultants, in terms of qualification, impartiality and deontology? (E)

Yes, consistent with DFID's procurement policy.

4. Evaluation Partnerships and Capacity-building

4.1 To what extent are beneficiaries involved in the evaluation process? (E)

To the extent that programmes' monitoring systems include this, sometimes supplemented by evaluation case studies. To date Country Programme Evaluations have not directly involved beneficiaries. The next round of CPE's will pilot a methodology for this consistent with the resources available for CPEs.

4.2 To what extent does the agency rely on local evaluators or, when not possible, on third party evaluators from partner countries? (E)

Standard requirement to include local consultants in all EVD studies.

4.3 Does the agency engage in partner-led evaluations? (E)

Yes, e.g. supported Uganda's evaluation of its Poverty Eradication Action Plan; Paris Declaration Evaluation.

4.4 Does the unit support training and capacity building programs in partner countries? (E).

Yes e.g. through World Bank, IDEAS, IPDET, Regional evaluation networks, and impact evaluation networks and foundations; sponsoring attendance at training events, workshops and conferences.

4.5 How do partners/beneficiaries/local NGOs perceive the evaluation processes and products promoted by the agency/country examined (in terms of quality, independence, objectivity, usefulness and partnership orientation)? (OS)

5. Quality

5.1 How does the evaluation unit ensure the quality of evaluation (including reports and process)? (E)

Through its formal Quality Assurance process, currently being strengthened and systematized.

5.2 Does the agency have guidelines for the conduct of evaluation, and are these used by relevant stakeholders? (E)

Yes.

5.3 Has the agency developed/adopted standards/benchmarks to assess and improve the quality of its evaluation reports? (E)

Yes, based on a combination of the DAC standards and those for engagement in humanitarian situations and fragility.

5.4 How is the quality of evaluation products / processes perceived throughout the agency? (M)

We do not have systematic evidence on perception of evaluation products. My impression is that reports are seen as professional and thorough.

There is an appetite for greater use of quantitative evidence in evaluations – this relates directly to the greater setting of quantitative targets with baselines and means of verification in our planning processes, on which DFID is now making some progress.

There is perhaps a question about whether the style of presentation has kept pace with the more accessible communications styles now used for some of our other products.

6. Planning, Coordination and Harmonisation

6.1 Does the agency have a multi-year evaluation plan, describing future evaluations according to a defined timetable? (E)

Yes.

6.2 How is the evaluation plan developed? Who, within the aid agency, identifies the priorities and how? (E, M)

EVD staff develop a draft 3 year topic list through internal and public consultation and analysis of external documentation including the DAC's new database on evaluations.

The evaluation plan is developed through wide consultation across DFID, with external stakeholders and with IACDI. The commitment to a rolling programme of country programme evaluations and to cover all main policies sets the parameters for this.

6.3 In DAC Members where ODA responsibility is shared among two or more agencies, how is the evaluation function organized? (N/A)

In broad consistency with the DAC's guidelines for conduct of joint evaluations e.g. international reference groups, management groups involving donor and partner countries etc.

6.4 Does the evaluation unit coordinate its evaluation activities with other donors? (E)

Yes through the DAC Evalnet.

6.5 How are field level evaluation activities coordinated? Is authority for evaluation centralized or decentralized? (E)

Coordination in the field for joint evaluations is managed through the management group. Overall authority is vested centrally through the new Evaluation Policy.

6.6 Does the evaluation unit engage in joint/multi donor evaluations? (E)

Yes, wherever possible in line with Paris and Accra commitments.

6.7 Does the evaluation unit/aid agency make use of evaluative information coming from other donor organizations? (E)

Yes, in scoping topics for evaluation, and in developing study approach papers.

6.8 In what way does the agency assess the effectiveness of its contributions to multilateral organizations? To what extent does it rely on the evaluation systems of multilateral agencies? (E, M)

The new policy proposes full reliance on evaluation systems of multilaterals, and periodic independent evaluation of the effectiveness of DFID's partnership with each agency.

In terms of assessing the effectiveness of spending investments by multilateral agencies, DFID largely relies on the evaluation systems of those agencies. Our Evaluation Department has varying degrees of contact with the evaluation functions of these agencies.

DFID has a number of approaches for assessing the effectiveness of its contributions to multilateral agencies, which contribute to its decisions about how to allocate resources between agencies. For example, DFID publishes institutional strategy papers setting out the main objectives of its contributions to multilateral agencies – progress is monitored by operational teams. DFID has worked with other shareholders to develop approaches to assessing the effectiveness of multilateral agencies' own systems (including MOPAN).

7. Dissemination, feedback, knowledge management and learning

7.1 How are evaluation findings disseminated? In addition to reports, are other communication tools used (press releases, press conferences, abstracts, annual reports providing a synthesis of findings)? (E)

Yes to all, and in addition through Head of EVD's Annual Report, specialist adviser networks, aid effectiveness networks and seminars.

7.2 What are the mechanisms in place to ensure feedback of evaluation results to policy makers, operational staff and the general public? (E)

Director-level responsibility is assigned for management response for all evaluations; all reports are placed on DFID's website. Head of Evaluation sits on Development Committee that approves and monitors DFID policies; Country Programme Review Committee; Investment Committee; Management Board annual discussion on evaluation. Dissemination of relevant findings at specialist adviser retreats, and through network newsletters - especially the Aid Effectiveness newsletter.

7.3 What mechanisms are in place to ensure that knowledge from evaluation is accessible to staff and relevant stakeholders? (E)

Covered in above answers – use of internal and external websites, policy briefs etc.

7.4 Is evaluation considered a 'learning tool' by agency staff? (E, M)

For management to respond and see also IACDI/NAO survey of stakeholder views on DFID evaluation (referenced in TORs for this quality review).

Yes. The process of evaluation is an important process of reflection for operational teams. Evaluation products are considered an important part of the evidence base for development effectiveness; they are a key source for the knowledge base of our professional advisers, and an important resource for development of new policy and programming.

8. Evaluation Use

8.1 Who are the main users of evaluations within and outside the aid agency? (E, M, OS)

See survey as above.

Policy and operational staff within DFID are important users of our evaluations. We do not have systematic evidence on wider use; but I would expect that joint evaluations, and maybe also DFID's own evaluations, are similarly used by other donors.

Civil society consultees showed very little knowledge of DFID evaluation products, one drawing an explicit contrast with those of the World Bank's Independent Evaluation Group, which they followed with some attention.

8.2 Does evaluation respond to the information needs expressed by parliament, audit office, government, the public? (E, M, OS)

For management to respond.

Evaluations are not one of the primary ways in which DFID responds to the information needs of parliament, other parts of government or the public. The National Audit Office draws heavily on evaluation studies where relevant. The audience for evaluation material is more within DFID and the development community.

CSOs consulted did not, at present, give much attention to DFID evaluation reports (NB this is based on consultation of only a few groups known to be interested in aid matters: it is quite possible that groups concerned with individual topics, such as gender of HIV/AIDS, would have taken a different view).

8.3 Are there systems in place to ensure the follow up and implementation of evaluation findings and recommendations? (E)

Yes as explained above.

8.4 How does the aid agency/ministry promote follow up on findings from relevant stakeholders (through e.g. steering groups, advisory panels, sounding boards)? (E,)

EVD has appropriate steering groups and reference panels for all its evaluations; management to respond on operational and policy follow up.

8.5 Are links with decision making processes ensured to promote the use of evaluation in policy formulation? (M)

Yes – the follow-through from evaluation to policy formulation is strong (DFID's recent policies on HIV/AIDS, trade, gender, budget support all draw directly from evaluation studies).

8.6 Are there recent examples of major operation and policy changes sparked by evaluation findings and recommendations? (M)

DFID's action plans on gender and results were driven in large part by evaluation evidence on weaknesses in these areas. DFID's evolving approaches to HIV/AIDS, and budget support were significantly influenced by evaluation findings.

8.7 Are there examples of how evaluation serves as an accountability mechanism? (M)

Evaluations are not part of formal accountability mechanisms within DFID. But they are used informally by Directors and Director-Generals to hold country programme managers and policy managers to account. Managers certainly see evaluations as some sort of verdict on their personal performance.

8.8 What are the perceptions of non evaluation actors (operation and policy departments, field offices, etc) regarding the usefulness and influence of evaluation? (M)

We do not have systematic evidence on perceptions of EVD's work. My impression is that evaluations are seen as useful and influential on shaping future organisational thinking.

They are also often seen as useful and challenging to the teams whose work is evaluated – although they sometimes have the sense that the evaluation is telling them what they knew already and feeding back evidence that they have given to the evaluation team. Stronger use of external evidence sources will help with this.

References:

- *DAC Principles on Evaluation of Development Assistance [OCDE/GD(91)208]*
- *Review of the DAC Principles on Evaluation of Development Assistance OECD(1998)*
- *Strengthening Evaluation Systems in Member Agencies [DCD/DAC/EV(2004)1]*
- *Evaluation systems in DAC Member Agencies; a study based on DAC Peer Reviews, presented at the Second Meeting of the DAC Network on Development Evaluation , Paris 9-10 2004*

ANNEX 5: EUROPEAN COMMISSION

1) **Is the operating environment favourable to high-quality evaluations?** [DAC Questionnaire Sections 1, Policy; 2, Impartiality, Transparency and Independence; and 3.1-3.4, Resources and Staff]

a) Policy

A single evaluation service covers the three External Relations Directorates which between them manage the development aid programmes of the European Commission, DG RELEX, DG DEV and DG AIDCO (EuropeAid). Separate evaluation services exist for other Directorates, including DG ECHO (emergency and humanitarian aid).

The overall policy is set out in a key paragraph of the Commission's decision of May 2000 on reform of the management of its aid programmes. This makes an important distinction between the evaluation of individual projects (seen as a function integral to the planning cycle and so a core task of AIDCO) and the evaluation of the results of regional and sectoral policies, programmes and programming performance, which are seen as crucial to the success of external policies as a whole and where, to secure objectivity, independent evaluation is considered necessary.

A little more detail is provided in five paragraphs of an Inter-Service Agreement of June 2001. This repeats that AIDCO, not the evaluation service, is itself responsible for the evaluation of individual projects and also for sending a copy of each completed evaluation report to a central database (not maintained by the Evaluation Unit). Indeed, under the 'Results-oriented Management' system of AIDCO, every project over € 1 Million is supposed to be assessed by independent consultants annually, and for 30% of them after project closure. This means, in principle, 1,400 reports a year at the project level.

In the same Agreement, the evaluation service is duly charged with the task of evaluation the results of Country/Regional and sectoral policies, programmes and programming performance. This means that the core task of the evaluation service is to examine the impact of policies and programmes, as distinct from projects. Further provisions, noted below, cover the evaluation programme, the definition of terms of reference, and reports on evaluation findings.

b) Management and Coverage

The evaluation service is located in DG AIDCO. Its coverage is in principle any policies and programmes carried out by the three Directorates-General.

There is complete separation between evaluation and internal audit in the Commission. The Court of Auditors is however increasingly seeking to assess the results achieved by EC spending, and this is producing more contacts between the Court and the evaluation service, the Court preferring on the whole to investigate after than before an evaluation report.

c) Independence

The Head of the evaluation service reports to the two Commissioners who between them are responsible for the three Directorates-General, in the present Commission Benita Ferrero-Waldner (RELEX and AIDCO) and Louis Michel (DEV). The Head of the service is a career Commission official.

d) Resources

The budget of the service is some E6 million (say £5 million) a year. Given that its remit does not include the evaluation of any specific projects, it is difficult to relate this to overall EC Development spending (some E9 billion a year) in any very close way. The Service itself considers that its coverage amounts to about half of the activities within its mandate. It has 11 professional staff. Its output is about 17 reports a year.

e) Staff competencies

Of present staff, beside the Head of Service 2 other persons came to their position from a background in evaluation. All new professional staff receive training, both formal training and about 6 months' 'on the job' before themselves being required to manage an evaluation.

2) **Is the way in which the topics for evaluations are decided likely to encourage relevance and buy-in?** [DAC Questionnaire 6.1-6.3, planning and priority setting; 6.8, approach to multilateral aid]

Under the terms of the Inter-Service Agreement, a draft programme of evaluation exercises – sectoral, thematic, country, regional, as well as the 'policy mix' and programming – is to be prepared each year by the evaluation service, examined by the three Directors-General and by them to the Board of AIDCO (Chaired by the Commissioner for RELEX and AIDCO) for approval. In addition to this, the Head of Service proposed a comprehensive plan for evaluation covering the whole 7 years of the current Financial Perspectives (2007-2013). This was scaled down to fit existing resources, and signed off by Ms Ferrero-Waldner, and provides a basis for the annual planning exercise.

The service has evaluated EC grants to multilateral bodies, including a recent comparison of EC grants to the World Bank and the European Investment Bank (the EIB's own evaluation service is of course responsible for evaluating EIB's use of its own resources).

3) **Is the approach to involving partners and other donors in evaluations logical and reasoned?** [DAC Questionnaire 4.1-4.3, involvement of partner countries; 6.4-6.7 involvement of other donors]

The Service does not have any target for the proportion of joint evaluations, and has a lively sense of the difficulty of putting them together, both practically (transferring money from or to other member states raises near-insuperable difficulties)and in terms of reconciling different approaches (a nice example of the latter being an African country where this approach resulted in not evaluating the Commission's huge budget support programme but evaluating several small projects of relatively minor bilateral agencies). Nevertheless, the EC has launched a new methodological approach for evaluating budget support that can only be done jointly with other donors and partner country.

The evaluation service has no mandate to help build local evaluation capacity in recipient countries (and indeed has some doubts as to how many such countries are yet ready for a serious investment in this area). It does however encourage evaluation consultants to make use of Southern expertise, though this is not a formal requirement of contract award.

4) **Are evaluations managed with a view to a quality product?** [DAC Questionnaire 3.5, selection of consultants; 5.1-5.3 managing for quality]

The Head of service sees quality as essential to the credibility of the evaluation process, and hence has put in place a number of procedures to this end. The key elements are set out in a set of short papers which together form the "Evaluation Methodology for European Commission's External Assistance". The service operates a 'quality grid' of nine criteria, each of which is marked on a five point scale, which enables them to track quality of reports in a systematic way. The latest survey of 48 evaluations commissioned by the service shows that 44 were rated as 'good', suggesting that a consistent quality level is being achieved on this measure (only one study fell below 'good' and only three rated as 'very good').

As in most bilateral agencies, studies are carried out by external consultants.

Key elements in ensuring quality are: careful attention to terms of reference by the evaluation service at the outset; agreeing with the selected consultants a list of no more than 10 evaluation questions, with explicit 'judgement criteria' for each question, and strong attention to how the consultants' report is written. The evaluation questions do not replicate the DAC criteria but should relate to them.

During the evaluation cycle, the roles of the evaluation service and the consultants are supposed to be clear: thus there is shared responsibility for the evaluation questions and for communication, but the evaluation service is exclusively responsible for terms of reference and the consultants for the evaluation judgements (though the evaluation service may critique their methodology) and for the final report.

Studies will normally include an outside evaluators' team and have a 'reference group' of between 8 and 15 people chaired by the evaluation service, and including representatives from the Commission (country delegation, country desk, thematic staff) and, for country evaluations, the relevant Ambassador in Brussels.

While the Head of Service is in principle supportive of experimental design approaches, he considers them to be of marginal utility for the questions that he is required to evaluate.

5) **Are findings disseminated in a way likely to secure impact?** [DAC Questionnaire 7.1, 7.3 and 8.4]

A 1-2 page summary of each evaluation report is sent to the two Commissioners, and the full report goes to the reference group and to other interested internal stakeholders (typically about 50 people). Copies are also sent to the Court of Auditors, the European Parliament, and the Council of Ministers and to the evaluation units of all member states. Reports are routinely published on the Commission's website, and are summarised in a few lines in AIDCOs annual report. (A separate annual report by the Head of service is no longer published.)

A public seminar will be held to discuss each completed report, and for country reports this will be held in the country concerned (at the stage of the draft final report, so that stakeholders can react).

The evaluation service considers that the internal impact of its work is increasing sharply.

6) **Are recommendations followed up systematically?** [DAC Questionnaire 7.2 and 8.3]

The practical vehicle for following up evaluation recommendations is the 'fiche contradictoire', in which the Services that are covering the subject of the evaluation have to state whether they agree, partially disagree or completely disagree with each recommendation. One year later, a review of any action taken is made by the evaluation service on the basis of the information given by the concerned services and the results of this review are published on the Commission's website. A proposal to require an 'Action Plan' on follow-up to recommendations was not accepted.

7) **How are evaluation products perceived by outside stakeholders?** [DAC Questionnaire 4.5, local stakeholders; 5.4 and 7.4 development agency staff; 8.1-8.2 other stakeholders, including Parliament and public]

As usual, this is hard to judge without more work than this assignment provided. The view of the Head of the evaluation service is that Civil Society organisations make significant use of EC evaluations, not least given their attention to impact. The European Parliament, by contrast, has paid them no significant attention.

8) **What role does the evaluation department play in relation to evaluative material commissioned by other parts of the agency?** [Not covered in DAC Questionnaire]

There is no formal requirement on the evaluation service to support or assess 'decentralised' evaluation. However, the service did recently assess the quality of 40 such evaluations, chosen in a randomised manner, and using a slightly

simplified version of the 'quality grid' mentioned above. Its conclusion, very much in line with Swedish and Dutch studies, was that half scored 'good' or above and half scored 'poor' or below. Studies were judged relatively positively on relevance and as having credible findings while, paradoxically, scoring less well on 'defensible design'. The service does also provide informally guidance to operating units which ask for it.

ANNEX 6: IRISH AID

1) Is the operating environment favourable to high-quality evaluations? [DAC Questionnaire Sections 1, Policy; 2, Impartiality, Transparency and Independence; and 3.1-3.4, Resources and Staff]

a) Policy

Irish Aid operates under an evaluation policy published in December 2007. In this, evaluation is defined as the systematic and objective assessment of the design, implementation and results of an on-going or completed project, programme or policy by assessing the effectiveness of the intervention against its stated objectives. Learning and Accountability are stated as the two objectives of evaluation. The policy states that while both objectives must be satisfied in evaluation activities, individual evaluations need not always fulfil these objectives to the same degree and that the main priority of each planned evaluation should be considered at the outset.

The DAC Principles and Criteria are a central point of reference in the evaluation policy. Gender, environment, HIV/AIDS and governance are considered as cross-cutting issues.

b) Management and coverage

Within Irish Aid, evaluation is managed by the Evaluation and Audit Unit. Its Mission is:

"To maintain an efficient, effective, relevant and independent evaluation and audit function within Irish Aid through the execution of evaluation and audit exercises, facilitating acceptance of their findings and contributing to policy development within the overall programme."

Irish Aid is unusual in combining evaluation and internal audit functions in a single unit. The head of Unit considers that this arrangement brings valuable synergies, and there are no plans to change it.

The Unit's mandate covers evaluation of Official Development Assistance funds administered by the Department of Foreign Affairs but not expenditure administered by other Departments. In respect of multilateral institutions, reliance is placed on the multilaterals' own evaluation systems and on joint evaluations where possible.

c) Independence

In respect of evaluations, the Unit reports to the Director General responsible for Irish Aid and to a fully independent Audit Committee, whose membership excludes staff of the Department of Foreign Affairs (though the Committee merely informs itself of evaluation reports). In order to enhance

the independence of the evaluation arrangements for the Irish Aid programme, the Advisory Board for Irish Aid (ABIA)[16] has the right to see, and to comment on, the reports produced by the Evaluation and Audit Unit.

In practice, while there can be robust discussions, particularly about tone and language, with operating units, Unit staff do not recall any senior management pressure to adjust substantive findings.

d) Resources

There is no separate budget line for evaluation, or indeed for the Unit as a whole, which forms part of a wider administrative budget. It has however been given a degree of priority, and is increasing its full-time evaluation positions from three to four staff. It delivers 3-4 reports a year, designed to cover strategic evaluations across the entire development programme. Expenditure in the last fiscal year was around E500,000.

e) Staff Competencies

Irish Aid evaluation staff are drawn from the body of staff of the Department. The present Head of the Unit has worked in evaluation before, but this would not be typical. Training is essentially 'on the job'.

2) Is the way in which the topics for evaluations are decided likely to encourage relevance and buy-in? [DAC Questionnaire 6.1-6.3, planning and priority setting; 6.8, approach to multilateral aid]

The policy document states that the decision to evaluate should be informed by the following criteria:

- Policy Relevance – what is the relevance of the development intervention in the context of Irish Aid's strategies, policies and goals?
- Usefulness – what use will the outputs of the evaluation be to the interests, plans and priorities of Irish Aid, its partners and other stakeholders?
- Accountability – was the intervention carried out in an efficient and cost effective way?
- Innovative value and replicability – does the development activity proposed for evaluation represent a new and innovative way of dealing with development challenges? Is it likely that it can be replicated and adopted on a wider scale?
- Financial importance – what is the importance of the development activities proposed for evaluation in terms of the scale of funding?

[16] Membership of ABIA however lapsed in December 2008, so it is not functioning at this point. It is understood that the Irish Government is considering what future arrangements should be put in place. The ABIA has commissioned its own evaluations in the past, for example on *Measuring Impact: the Global and Irish Aid Programme*.

- Risk – would a failure of the development intervention pose a particular threat to the organisation's capacity to achieve its goals? (Amongst the risk factors that should be taken into account are those of a political, economic, financial or organisational nature).
- Evaluation feasibility – is it possible to carry out the evaluation as intended? Are there reliable data, competencies and financial resources etc available?
- Cost-benefit – is the evaluation likely to produce results that can justify its costs?

Armed with these criteria, the Evaluation and Audit Unit puts together (a) a three-year rolling work plan, and (b) an annual operational work plan which is approved by the Senior Management Group of Irish Aid, as well as the Audit Committee. This is done in a consultative way, and the programme may reflect specific requests from the field (eg a recent report on Tigray, Ethiopia) or areas where management has a specific interest (eg the latest evaluation of Irish Aid to Uganda was put in place with a view to discussion of future policy; the evaluation of aid to East Timor was closely related to issues needing management review). The Senior Management Committee does not however necessarily accept all the proposals by the Unit.

3) **Is the approach to involving partners and other donors in evaluations logical and reasoned?** [DAC Questionnaire 4.1-4.3, involvement of partner countries; 6.4-6.7 involvement of other donors]

Ireland has no quantified policy on the proportion of its evaluations that should be joint, but is very open to such approaches where relevant. The policy statement says that Irish Aid will actively seek to undertake joint evaluations when and where appropriate. (For example, Irish Aid would see joint evaluation as the norm for a sector in a particular country where the sector is co-funded by a group of donors, or for major cross-country studies; but there are cases where the evaluation is specific to the Irish programme, as with E Timor and Uganda, where a joint approach would be inappropriate.)

As for use of local partners, Ireland uses local consultants where possible, though not so far in lead roles. The Unit welcomes in principle the concept of recipient-led evaluation, and was a partner in the recent Government of Uganda-led Evaluation of the Poverty Eradication Action Plan. But it has no mandate to support local capacity-building.

4) **Are evaluations managed with a view to a quality product?** [DAC Questionnaire 3.5, selection of consultants; 5.1-5.3 managing for quality]

The Unit attempts to build in quality through a number of methods.

In accordance with the policy statement, Irish Aid has adopted the Swedish International Development Agency Evaluation Manual 'Looking back, moving forward' as its core reference document for the management of evaluation assignments. It uses the DAC quality guidelines.

All evaluations are managed by the Unit and supervised by a steering committee or reference group, which will always involve at least one stakeholders from outside Irish Aid (typically from another Irish institution or experienced individual), as well as representatives of the unit being examined and of the evaluation unit itself. This group will examine the Terms of Reference at the outset (they will also be shown to local stakeholders), and follow the various stages of the evaluation.

Evaluations are normally carried out by teams of consultants, but an interesting model being tried for an evaluation of Ireland's humanitarian contributions to Liberia and Sierra Leone is to hire a consultant team leader, with the members of the team coming from the evaluation unit itself, from a staff member of the humanitarian unit not involved in either country, a member of the Diplomatic Service with knowledge of the area and a local consultant.

5) Are findings disseminated in a way likely to secure impact? [DAC Questionnaire 7.1, 7.3 and 8.4]

The policy statement requires that a systematic approach and plan for dissemination of the report, its findings, conclusions and main lessons, should be established as part of the ordinary planning process of an evaluation. Dissemination processes include formal and informal workshops and briefings. Within the Department, all studies are now made available through the new Intranet. There is no system of circulating summaries, but a group looking at knowledge management in the Department in looking at what more can be done to promote lesson learning.

Reports are also normally placed on the Department's web-site, though this was not fully up to date. A Press Release would normally also be put out on publication. So far, these have led to occasional newspaper reports, but in general public and Parliamentary attention to the reports has been limited (in contrast to audit reports, which typically get more coverage).

6) Are recommendations followed up systematically? [DAC Questionnaire 7.2 and 8.3]

A Management response is required to all reports and is usually provided before publication of the report. When recommendations are rejected by the operational department / unit concerned, the reasons for rejection should be clearly stated .The formal system of monitoring follow-up to value for money audit reports is being extended to evaluations, providing further systematisation of responses. Management responses are also seen by the Audit Committee.

The 'Coherence Group' of senior officials considers the recommendations of all evaluation reports from the Unit. A further check is that submissions for new spending commitments to the Programme Appraisal and Evaluation Group (which includes members of other government departments and a few high-level outside experts) have to certify that the Unit has been consulted.

7) How are evaluation products perceived by outside stakeholders? [DAC Questionnaire 4.5, local stakeholders; 5.4 and 7.4 development agency staff; 8.1-8.2 other stakeholders, including Parliament and public]

Encouragingly, there has a good deal of positive interest internally in several evaluations, the East Timor evaluation being a good example. As noted above, external interest, even among NGOs, has to date been somewhat limited.

8) What role does the evaluation department play in relation to evaluative material commissioned by other parts of the agency? [Not covered in DAC Questionnaire]

The evaluation policy requires the Unit to provide advice, training and support to monitoring and evaluation work of the operational departments and, where requested, Irish non-government organisations working in development. The Unit spends about 20% of its time assisting operating units with their own evaluation work, Southern Sudan being a current example. The Unit's business plan attempts to capture what is known about decentralised evaluative work in Irish Aid.

ANNEX 7: NORAD EVALUATION DEPARTMENT

1) Is the operating environment favourable to high-quality evaluations? [DAC Questionnaire Sections 1, Policy; 2, Impartiality, Transparency and Independence; and 3.1-3.4, Resources and Staff]

a) Policy

NORAD was radically restructured in 2004, from an organisation which delivered most Norwegian bilateral aid to one responsible for technical advice, research, quality assurance and evaluation and for programmes with civil society and the private sector, and where the learning function was given high priority.

As part of this restructuring, NORAD's Evaluation Department now has the central role in all aspects of evaluation of Norway's development cooperation policies and programmes. It does so under the terms of a mandate from the Ministry of Foreign Affairs and its own Evaluation Policy covering the period 2006-10.

This states that the two main purposes of evaluation of Norwegian development cooperation are to:

a) Systematise knowledge of results and performance which can help improve similar activities in the future (learning function)
b) Evaluate whether an activity has been performed as intended and/or whether the expected results have been achieved (control function, documentation).

The policy states that Evaluations need to be of high quality and *objective, verifiable, transparent, impartial and independent*. The DAC criteria are given specific mention, and any departure from them in a particular study needs to be acknowledged and defended.

In the period 2006-10, the Department is required to address 6 key issues:

1. Quality assurance of *all* development cooperation
2. Stronger focus on *results* of Norwegian aid
3. Adapt evaluation work to *new aid modalities*
4. Improved *communication* of evaluation results and improved learning
5. Strengthen evaluation as basis for *policy development*
6. Strengthen *quality* and reliability of evaluation activities

b) Management and Coverage

A separate Evaluation Department in NORAD is charged with implementing the policy. Its coverage is any activities funded under Norway's development cooperation budget. In respect of multilateral aid, it however covers only earmarked contributions, not Norway's core contributions. (The latter are not

excluded per se, but it is not seen as realistic for a single bilateral donor to conduct an evaluation of the effectiveness of particular multilateral agencies.) It has also looked at policy coherence.

c) Independence

The Head of the Evaluation Department formally reports to the Secretary-General of the MFA (the highest civil servant in the Ministry). In practice, the Secretary-General has delegated this role to the Deputy Secretary-General (the most senior official with full-time responsibility for development cooperation within the Ministry). The Head of Evaluation Department submits papers through the Head of NORAD, who may comment on them but not amend them. The Head of Evaluation also publishes an Annual Report in his own name. (Separately, the NORAD Director also produces an annual report on the Norwegian Aid Programme, which is acquiring significant profile.)

d) Resources

The Department's budget is of the order of NKr 22 million a year (around £3 million), which is around 0.14% of Norway's bilateral aid (including multi-bi commitments, as opposed to core multilateral contributions). It has 10 professional staff.

e) Staff competencies

About half the professional staff are career NORAD staff and half have been recruited since 2004 specifically for their evaluation skills and experience, mostly with Doctorate-level qualifications. NORAD have found a lot of interest when it has advertised positions externally.

2) Is the way in which the topics for evaluations are decided likely to encourage relevance and buy-in? [DAC Questionnaire 6.1-6.3, planning and priority setting; 6.8, approach to multilateral aid]

The policy document sets out four criteria for selection of evaluation topics: the characteristics of the operation (in principle, more complex interventions would score higher on this), risk, significance (which includes size, potential impact etc), and cost-benefit.

The Evaluation Department works to a 2-3 year rolling plan, updated annually. It consults widely on what should be included, but the head of the Evaluation Department has the last word on the content of the programme (subject to the caveat that the Minister, or NORAD's Director, may direct it to undertake a particular evaluation: the Evaluation Department has in practice responded on a couple of occasions to requests from the Ministry). It seeks to cover all main areas of expenditure on a 3-4 year cycle.

NORAD seeks to be responsive to well-founded suggestions from operating units. For example, it proposed an evaluation of Norway's country programme to Zambia. The Norwegian Embassy in Lusaka considered that such an evaluation would not be particularly useful unless set in the context of why it

was that average income per head in Zambia was still not significantly higher than at Independence. The result of this dialogue was that the study went ahead on a much more ambitious basis than first planned, involving three field studies of key sectors where Norway had had interventions over a long period, and an assessment of Zambia against the 'Neo-Patrimonial State' model. The upshot has been a report which has raised a number of difficult issues, both in Zambia and in Norway, but which is increasingly seen as a valuable reference document with implications beyond Zambia.

3) **Is the approach to involving partners and other donors in evaluations logical and reasoned?** [DAC Questionnaire 4.1-4.3, involvement of partner countries; 6.4-6.7 involvement of other donors]

Norway has a clear policy of a broadly 50:50 division between joint evaluations and those carried out by NORAD for its own purposes.

In respect of involvement of partners in beneficiary countries, the policy document asks the Department to 'encourage evaluations spearheaded by a partner organisation in a partner country where the necessary expertise is already in place or can be built up during the course of the evaluation'. Two recent examples illustrate how this works in practice.

In Uganda, NORAD decided that an evaluation of the activities of Norwegian NGOs in Northern Uganda would be tendered locally. It was won by a local company (Image Consult Ltd). NORAD found that they needed to devote significant resources to managing the study and to ensuring quality, but the final product appears to have many useful insights not just into the direct impacts of what was delivered but on issues such as sustainability, changing perceptions, and adapting aid to a changing local security environment.

In Nepal, the Norwegian Embassy responded to a NORAD enquiry about possible priorities for evaluation by proposing that Norway undertake on behalf of the Ministry of Education and Sport an evaluation of a the country's Education for All Sector Programme, 2004-09. Such an evaluation had been foreseen in the financial agreement between the Government of Nepal and the donors at the outset of the Sector Programme. NORAD agreed to finance and manage the evaluation, which was carried out by a team of international and Nepali consultants, with the Nepali Ministry as the client.

Despite these examples, NORAD concede that partner involvement remains quite limited. Ministries are routinely consulted on evaluations but seldom feel much ownership of the product. In the case of the Zambia study, a formal response was sent by the Zambia Government, but this was not typical.

4) **Are evaluations managed with a view to a quality product?** [DAC Questionnaire 3.5, selection of consultants; 5.1-5.3 managing for quality]

Norway's normal policy is to tender internationally all major evaluation studies (the Uganda example quoted above was a relatively modest-sized one). Their experience is that, nevertheless, there is a wide range of quality, and they undertake a good deal of hands-on management. Issues of concern will normally

include whether the evaluation questions have been answered, and whether the findings reflect the analysis. NORAD will normally deploy 2 staff members to supervise and manage each major contract, a manager and a partner who can provide advice.

In addition they will normally employ one or more External Advisers for each major study, these usually being evaluation specialists, and whose role is essentially to advise NORAD on issues related to evaluation quality.

NORAD have tried both detailed and more 'open' Terms of Reference for evaluation studies (a recent, and good-quality, evaluation of their fisheries projects being an example of the latter).

The Evaluation Department uses the DAC quality standards.

All stakeholders are consulted on the Terms of Reference, on the Inception Report and on the first draft of the Final Report. (In the latter two cases, NORAD will have satisfied itself on the basic quality of the work before distributing the document.) NORAD will send the consultants copies of all comments, but the only ones that they have to take into account are those of NORAD itself. NORAD consider that ***it is important that the stakeholders' groups are not regarded as steering groups***. The Evaluation Department would however always respect the consultants' right to decide on findings, conclusions or recommendations, and would not try to dictate or steer the consultants in other issues than quality. The final draft is cleared more briefly. NORAD is happy for dissenting views to be recorded.

5) **Are findings disseminated in a way likely to secure impact?** [DAC Questionnaire 7.1, 7.3 and 8.4]

All reports are published, and a launch event will normally be held at the time of publication. In addition, the Annual Report of the Head of Evaluation contains both overview lessons from the year's evaluations, and a summary of each report published in the preceding year.

Key stakeholders will have been associated with the evaluation since the outset. More generally, within the Ministry of Foreign Affairs, a memorandum is sent to the Secretary-General of the Ministry a few weeks after publication giving a summary of the report and comments on it, and NORAD 's view of what action should be taken on any recommendations.

6) **Are recommendations followed up systematically? [**DAC Questionnaire 7.2 and 8.3]

Once the Memorandum has been sent to the Ministry, the responsible Department has 6 weeks to draft and submit to senior management (in practice, the Director-General) a plan to respond to the report. A report on action taken then has to be sent, again to the Director-General, within 12 months. The action taken is publicised in the Head of Evaluation Department's Annual Report.

7) **How are evaluation products perceived by outside stakeholders?** [DAC Questionnaire 4.5, local stakeholders; 5.4 and 7.4 development agency staff; 8.1-8.2 other stakeholders, including Parliament and public]

There have been a few evaluations that have caught public attention, for example the Zambia evaluation mentioned above. These would however be the exception.

8) **What role does the evaluation department play in relation to evaluative material commissioned by other parts of the agency?** [Not covered in DAC Questionnaire]

NORAD's Evaluation Department has a clear policy of devoting about 20% of its resources to support decentralised evaluative work. NORAD itself is well set up to support such activities, as it contains 6 thematic departments which routinely commission research and carry out reviews in their area of interest. Such reviews will typically be carried out by one staff member of the NORAD department concerned, one independent international consultant and one local consultant. These reviews are generally 'light-touch' and carried out over a few weeks. They would not in general qualify as evaluations in the strict sense, but NORAD's Evaluation Department sometimes provides professional advice, and now also makes the completed reviews available on the Internet.

ANNEX 8: SADEV

1) Is the operating environment favourable to high-quality evaluations? [DAC Questionnaire Sections 1, Policy; 2, Impartiality, Transparency and Independence; and 3.1-3.4, Resources and Staff]

a) Policy

The concept of an independent evaluation agency in Sweden goes back to a decision by Parliament in 2003 on policy for global development. The agency opened in January 2006.

Sweden has a system of relatively strong and independent agencies governed by periodic instructions, which have to be given by government as a whole, not just by the Ministry concerned, and are therefore not lightly changed. Originally the Ordinance for SADEV stipulated that SADEV should do both analyses and evaluations, but as from early 2007 the Government removed the reference to analyses from the Ordinance. The current wording of the Ordinance dates back to February 2009.

In SADEV's case, the relevant Ordinance is extremely brief. It states simply that "the task of the Swedish Agency for Development Evaluation is to, independently or in cooperation with other actors :

1. *perform evaluations of the Swedish international development cooperation and of the reform cooperation in Eastern Europe, and;*

2. *promote the development of evaluation capacity in the partner countries within the framework of the agency's regular activities according to 1". It goes on to require that SADEV's agency's evaluations address both results and effects (ie impact) of the development cooperation and efficiency of the use of development aid funds. SADEV is also required to distribute information regarding the results of its evaluation to the Swedish Parliament, the Swedish Government, government offices and other stakeholders in Sweden and internationally.*

This gives a mandate which is both very broad, but also with a focus on results, impact and efficiency.

An annual letter of appropriation from the government (not available in English) provides the budget and gives the possibility of more detailed guidance. The letter of appropriation is very brief but makes clear that evaluation topics must be highly relevant for decisions on development co-operation. Within this framework SADEV has independence in which topics it chooses and how it conducts evaluations.

b) Management and coverage

SADEV is headed by a Director-General (the current one since September 2008 with a background including acting as head of international affairs at

Sweden's National Audit Office, and working for IFAD and for the Nordic Development Fund, as well as for the Swedish Ministry for Foreign Affairs). Coverage is the whole of Sweden's development cooperation. Within this area the Ordinance allows SADEV to do work for other clients. This enables SADEV to cover the activities of multilateral agencies which Sweden finances (eg a critical assessment of UNDP's country-level evaluations). It has some limits, for example in relation to looking at the coherence of Sweden's overall policies that may impact on developing countries unless there is a specific development cooperation-related angle that would fall within SADEV's brief.

c) Independence

SADEV is set up with probably the greatest measure of independence of any bilateral evaluation unit.

d) Resources

SADEV, which is still developing, has an annual budget of some SKr 20 million (some £2 million), and in total (permanent staff and short term experts) employs 15 evaluators. It produces about 7 evaluations a year.

e) Staff competencies

The agency's staff have qualifications mainly in the fields of economics, social and political science and law. At the outset most came from an academic research career, and Sida6 of the evaluators have a PhD. Some former consultants have been recruited and several have done assignments for development organizations. The person holding the new (since May 2009) position of Director of Evaluation has a strong background in evaluation rather than in development.

Training is predominantly on the job (and IPDET). It also includes seminars with external experts both in methodology and development cooperation.

The limited availability of competence in evaluation of development cooperation has been identified as a true challenge. In addition to promoting its core competencies (evaluation and development cooperation) SADEV tries to contribute to improved development cooperation evaluations by involving evaluation expertise also from other policy areas.

2) Is the way in which the topics for evaluations are decided likely to encourage relevance and buy-in? [DAC Questionnaire 6.1-6.3, planning and priority setting; 6.8, approach to multilateral aid]

As noted above, the breadth of SADEV's mandate and its independence gives the agency very considerable scope to determine its priorities.

In the beginning researchers brought their work into SADEV so as to give SADEV the possibility to within a short period of time publish a number of reports. Because of the more explicit emphasis on evaluations and relevance

SADEV nowadays stresses the importance of dialogue with stakeholders both before, during and after any evaluation - however paying attention to maintaining SADEV´s independent decisions. Evaluations in the development cooperation areas are given priority by government and parliament (democracy and human rights, environment and climate change, and gender equality and the role of women) as well as multilateral assistance. SADEV also tries to be cost-effective and therefore attempts to find synergies among evaluations. Staff are encouraged to put forward their own ideas for evaluation topics on one page. As of 2009, SADEV will put out for consultation draft proposals for evaluation topics and invite interested parties to come up with additional proposals.

SADEV and Sida now consult each other on their programmes. Some consultations are also held with the Swedish National Audit Office, the Swedish Agency for Public Management and the Research Service of the Parliament.

3) **Is the approach to involving partners and other donors in evaluations logical and reasoned?** [DAC Questionnaire 4.1-4.3, involvement of partner countries; 6.4-6.7 involvement of other donors]

SADEV has played a constructive role in taking forward thinking about the challenges and opportunities of joint evaluations, and carried out a survey in 2008 which has led to practical follow-up within the DAC Evaluation Network. It has participated in a joint evaluation in the field of anti-corruption. However, as SADEV does not operate by contracting its work out to consultants, its means of participating in a joint evaluation have to be either by participating in the management group or by taking responsibility for a discrete part of the joint study.

As noted, SADEV's brief involves capacity-building for partner countries, though specifically linked to its own mainstream evaluation activities. Experience is very limited to date, and has not proved trouble-free, but SADEV intends to make efforts in this area.

4) **Are evaluations managed with a view to a quality product?** [DAC Questionnaire 3.5, selection of consultants; 5.1-5.3 managing for quality]

SADEV's most distinctive approach to assuring quality is that all its evaluations are led by its own staff, rather than contracted out to consultants.

Late 2008 SADEV developed its own quality assurance instrument. This stresses the importance of good dialogue at the outset of any evaluation. There is collective discussion in seminar form within the agency of draft Terms of Reference for each new evaluation. Reports are subject to internal peer review, and looked at by external referees (who may be from abroad). The study team draft an implementation plan which is also scrutinised internally and by external referees. Plans will include specific dates for the presentation of emerging findings at key points, and draft reports go through a similar discussion process. hose whose work is being evaluated are also given the opportunity to comment.

5) Are findings disseminated in a way likely to secure impact? [DAC Questionnaire 7.1, 7.3 and 8.4]

The normal practice is to hold workshops with stakeholders, and for the Ministry of Foreign Affairs itself also to hold a workshop. Reports are publicised through press releases and articles, and are published on SADEV's website.

6) Are recommendations followed up systematically? [DAC Questionnaire 7.2 and 8.3]

As an autonomous separate agency, it is more difficult for SADEV than for internal evaluation entities to influence the follow up to its recommendations. Basically, it is up to the government what action, if any, they wish to take on a SADEV study. However, the new Director-General will make public in her Annual Report what actions have been taken, which will at least mean transparency on this score.

7) How are evaluation products perceived by outside stakeholders? [DAC Questionnaire 4.5, local stakeholders; 5.4 and 7.4 development agency staff; 8.1-8.2 other stakeholders, including Parliament and public]

So far, SADEV's reports have not had a great deal of public resonance in Sweden. One reason for this may be that they have been written in a somewhat academic style. One reason for amending the mandate to remove 'analyses' was recognition that there was a limited market for such products, but a stronger need for proper evaluations.

8) What role does the evaluation department play in relation to evaluative material commissioned by other parts of the agency? [Not covered in DAC Questionnaire]

Not directly relevant to SADEV.

ANNEX 9: SIDA EVALUATION DEPARTMENT(UTV)

1) Is the operating environment favourable to high-quality evaluations? [DAC Questionnaire Sections 1, Policy; 2, Impartiality, Transparency and Independence; and 3.1-3.4, Resources and Staff]

a) Policy

Sida underwent a major reorganisation in 2008 under new leadership. The existing evaluation policy document is out of date, and a new one is to be drafted. However, it is clear that the direction of policy will be to emphasise lesson-learning and closeness to the operational units of Sida. In part, this reflects the establishment of SADEV as a fully independent evaluation agency with a broader remit. A seminar between SADEV and UTV on how to complement each other's work was held in early 2009.

b) Management and Coverage

The evaluation function of Sida was from 1995-2008 part of a single department with internal audit. The previous head of this department came to the conclusion that the costs and burden of this arrangement outweighed its benefits, and UTV is now fully separate.

There is no senior committee with specific responsibility for evaluation, but a new 'Results Committee' is likely to be an important venue for discussion of evaluations.

The coverage of UTV is all Sida activities, which means in essence all Sweden's bilateral aid and multi-bi contributions.

c) Independence

The head of UTV, like his counterpart in audit, reports to the Director-General of Sida. The new policy document will define how 'critical edge' is to be maintained in the independent evaluation function while positioning UTV closer to operational staff.

d) Resources

The operating budget (ie excluding staff costs) is SKr 28 million a year (say E3 million), but of this SKr 10 million is allocated for capacity-building activities. This makes the operating budget a good deal smaller than the Norwegian equivalent for a programme of comparable size, but given the existence of SADEV this is logical. UTV produces something like 20 studies a year, including joint studies (see below), which is considerable for its size.

Staffing is made up of a head, 6 evaluation specialists, 1 publications expert and 2 assistants.

e) Staff competencies

Of the 6 evaluation specialists most have research backgrounds. One is from a Sida operations background.

2) **Is the way in which the topics for evaluations are decided likely to encourage relevance and buy-in?** [DAC Questionnaire 6.1-6.3, planning and priority setting; 6.8, approach to multilateral aid]

The system for deciding evaluation topics involves strong consultation across Sida to ensure relevance to agency concerns. UTV can and does add its own ideas. It also brings together ideas that are developing in different parts of the agency: for example it established recently that several regional teams were interested in the subject of land rights, and this will now be addressed in a multi-country evaluation.[UTV has no current programme of country evaluations. It did publish in 2007 a review of other donors experience, which concluded that such evaluations were perceived to be useful by the donors concerned in spite of their modest ambition in establishing a causal relationship between the support provided by development cooperation and development outcomes. The study however also noted the apparent weak interest shown by partner countries in such evaluations. UTV has no plans to embark on country programme evaluations.

3) **Is the approach to involving partners and other donors in evaluations logical and reasoned?** [DAC Questionnaire 4.1-4.3, involvement of partner countries; 6.4-6.7 involvement of other donors]

UTV does not have a target for the proportion of evaluations which should be done jointly with other donors, but in practice a high proportion of work is carried out jointly – in recent years as much as 80-90%. Indeed, UTV is now seeing its core evaluation work as in two parts, joint evaluations with other donors and Sida -only evaluations with joint ownership between UTV and the relevant Sida teams.

Sida has a clear policy of promoting local involvement and local capacity building, for which, as noted above, UTV has a specific budget of SKr 10 million (over E 1 million). It is looking at supporting the World Bank's initiative to promote regional resource centres in developing countries.

4) **Are evaluations managed with a view to a quality product?** [DAC Questionnaire 3.5, selection of consultants; 5.1-5.3 managing for quality]

UTV uses the model of contracting consultants or to Swedish universities for almost all its evaluations. It may undertake some more unit-specific studies (eg of what use is made of its evaluations) internally.

UTV places emphasis on getting the initial stages of evaluations right. An initial memorandum/approach paper is drafted and shared with all UTV's evaluators for comment. Terms of Reference are then drafted and discussed with a reference group of Sida stakeholders and, on occasion, also outside experts, and with UTV's Director.

A grid is used to track quality, based on the DAC Quality standards, which UTV finds work well.

UTV has no specific policy on promoting impact evaluations, but it does support 3IE.

5) **Are findings disseminated in a way likely to secure impact?** [DAC Questionnaire 7.1, 7.3 and 8.4]

UTV encourages discussion of emerging findings, and sees the reference group itself as an important vehicle for dissemination. Seminars/workshops will normally be held on completion, and a newsletter brings evaluation findings to all Sida staff. Reports are sent to all stakeholders, and are published on the Sida website and referred to in the Sida Annual Report.
However, UTV feels that more needs to be done to encourage lesson-learning – hence the directions referred to above.

6) **Are recommendations followed up systematically?** [DAC Questionnaire 7.2 and 8.3]

There is on paper a system for management response and follow up, but it has not been working effectively, partly due to a prolonged gap in the key staffing position on the control side of the agency. A special unit is now working on reinvigorating the system.

7) **How are evaluation products perceived by outside stakeholders?** [DAC Questionnaire 4.5, local stakeholders; 5.4 and 7.4 development agency staff; 8.1-8.2 other stakeholders, including Parliament and public]

Sida evaluation reports appear to have little public resonance in Sweden, partly because they are not put out in a pro-active manner. There is, for example, no specific link to Parliament.

8) **What role does the evaluation department play in relation to evaluative material commissioned by other parts of the agency?** [Not covered in DAC Questionnaire]

UTV published a study of decentralised evaluations ('Are Sida Evaluations Good Enough?') in early 2008, which was quite critical of the quality of such studies. Following that study, Sida has decided that such decentralised activities will in future be called 'reviews' rather than 'evaluations', a term which will be reserved for products subject to quality assurance by UTV (ie the core programme of joint reviews and UTV-led reviews described above). Nor will

they be published, as in the past, by Sida, except electronically. The focus will be on making use of these reviews in country. However, UTV will increase its assistance to decentralised 'Reviews' with a view to improving their quality, and will devote up to 30% of its resources to this task.

ANNEX 10: IOB (NETHERLANDS)

1) Is the operating environment favourable to high-quality evaluations? [DAC Questionnaire Sections 1, Policy; 2, Impartiality, Transparency and Independence; and 3.1-3.4, Resources and Staff]

a) Policy

IOB (in English, the 'Policy and Operations Evaluation Department') was established in 1977. In the major reorganisation of the Ministry of Foreign Affairs in the mid-1990s it was given the role of evaluating policies and programmes across the whole Ministry. There is no recent overall policy statement on evaluation policy, but there is a very clear operating mandate linked to the Ministry's objectives and described below. There is also a clear policy that the IOB will pay strong attention to impact.

b) Management and coverage

IOB is a separate Department led by a Director. It has significant links in two directions in particular:

1. *Along with the audit and inspection functions, it makes up the 'Control Cluster' of the Ministry. The three Directors meet regularly, and may coordinate their respective work or combine forces to tackle particular issues. The Algemene Rekenkamer (the Dutch Audit Office), like other such offices, is taking an increasing interest in value-for-money studies.*

2. *IOB works closely with the Departments responsible for objective-setting and for management for results. The Department for Finance and Economics (FEZ) is responsible for the Ministry's relations with the Finance Ministry, for the annual budget process and for the monitoring of progress against the 'Policy Articles' and 'Operational Objectives' which are laid down for the Ministry. As part of its role, FEZ has recently commissioned external consultants (including Professor Frans Leeuw) to review the use made of IOB's evaluations. This report, which is looking at 90 evaluations undertaken between 2004 and 2007, 30 of them in depth, is due in July. The Department for Development Effectiveness (DEK) manages the results-based management system for the Directorate-General for Development Cooperation, and as part of its functions is responsible for responding to the recommendations of evaluations. In 2008 it prepared a note on various aspects of the evaluation system which is discussed further below. DEK also leads on the biennial 'Development Results Report' of which the third such report issues in mid-May 2009, and which draws on IOB and other work.*

c) Independence

The Head of IOB reports to Ministerial level (the Dutch MFA has 3 Ministers: the Foreign Minister, the Development Minister and the Under-Secretary for European Affairs). Reports are submitted via the Secretary-General of the Ministry, who may forward them with his/her own

comments, but not amend them. The Dutch Parliament takes a strong interest in the independence of the evaluation function, and all reports are sent to it by the relevant Minister together with their own policy response. Parliament recently successfully pressed for the Head of IOB to be given the last word in deciding what development activities should be evaluated each year (the Secretary-General of the Ministry has the last word on the content of the rest of the evaluation programme). And Ministers have very recently agreed to the establishment of an External Advisory Panel to advise the Minister on evaluations – a remit with some similarity to that of IACDI, including the concern for quality.

Overall, IOB has a degree of independence unusual in the Dutch system, though there is a similar arrangement in the Ministry of Justice.

d) Resources

The operating budget (for studies and for research assistants) is E3.5 million a year. The staff include the Head (a former Ministry spokesperson and Ambassador) and 14 'inspectors' (senior evaluators), who are on the administrative budget, and who are supported by 10 junior evaluators. The output of studies varies from year to year, but over a 5 year period (2004-2008), 19 development cooperation and 5 foreign policy studies were published.

e) Staff competencies

The senior staff are roughly one third long-term IOB staff, who are making a career of evaluation, one-third specialists (among them secondees from Universities) with an evaluation background who are hired from outside the ministry for a 4 year period, and one third foreign service staff. The two latter groups will normally serve for 4 years. IOB make use of the University of Ontario course for formal training, and provide on-the-job training as well.

2) Is the way in which the topics for evaluations are decided likely to encourage relevance and buy-in? [DAC Questionnaire 6.1-6.3, planning and priority setting; 6.8, approach to multilateral aid]

As noted above, the Head of IOB has the last word on what development issues to evaluate, but in practice the system is driven importantly from two directions.

First, under the Ministry's objective-setting system, there are some eight overarching objectives (for development, they include eg 'Increasing Individual and Social Development') which drive the Ministry's activities, each with several more specific operational objectives. For each of these latter (eg 'give all children, young people and adults equal access to high quality education'), an annual six-year rolling evaluation plan is submitted as part of the budgeting process. This shows planned policy reviews and impact assessments of progress towards the objective in question for the four following years, the two previous years' reviews and assessments being included to complete the picture. In the case quoted, the 2008 plan included a major review in 2011 to be fed by impact

assessments in Uganda and Zambia in 2008 and by further joint evaluations with other donors in 2009 and 2010. This is a highly strategic approach, which enables the Netherlands, unlike most of its counterparts, to organise impact evaluations which can then contribute to the assessment of progress against specific policy objectives. It provides an unusually long planning horizon, though still with scope for flexibility eg on the timing of individual studies.

The second direction is less structured and more opportunistic. Parliament can and does propose topics for evaluation, recent examples being a major study of the Netherlands' Africa policy, including not just aid but also diplomatic efforts and economic relations, and an evaluation on the effectiveness of spending on public support for development. While the Head of IOB could technically refuse such requests, it would be very unusual to do so.

In both cases, the system guarantees that there will be a significant audience for the findings of the studies.

The Netherlands has no current programme of country evaluations.

3) **Is the approach to involving partners and other donors in evaluations logical and reasoned?** [DAC Questionnaire 4.1-4.3, involvement of partner countries; 6.4-6.7 involvement of other donors]

The IOB has no target for joint evaluations, but regards them as desirable where donors are collectively pursuing joint objectives. DEK encourage joint evaluation (including recipient countries) as part of their general commitment to following up the Paris Declaration.

The IOB has no specific programme of capacity-building for developing countries. However, it is strongly committed to working with recipient governments from an early stage and to encouraging them to engage with other stakeholders (such as the Teacher's Union in the case of the Uganda Education Impact Study). This close working relationship itself helps to build capacity: in the Uganda case, the Ministry of Education has indeed set up its own internal evaluation cell with a brief to follow up on the evaluation of education impact started jointly with IOB (eg on the problem of teacher absenteeism highlighted in the Uganda report.

An example of a genuinely joint product was an evaluation of the Dutch mixed credit programme in China run jointly by IOB and the Chinese National Center for Science and Technology Evaluation. An impact evaluation of water projects in Yemen had a women's NGO as counterpart.

4) **Are evaluations managed with a view to a quality product?** [DAC Questionnaire 3.5, selection of consultants; 5.1-5.3 managing for quality]

IOB is unusual among bilateral evaluation agencies (except for SADEV) in its practice of having its own senior evaluators lead the teams for each evaluation, rather than act just as evaluation managers. Consultants and/or universities are usually brought in as part of the teams, but not as the leaders of evaluations. IOB

also claim that they can take a tougher line than some other bilateral agencies in rejecting sub-standard work by consultants and being unmoved by threats of legal action (which have not in fact materialised), because the experience of their team leaders enables them to take such a line.

New guidelines for evaluation are being developed, drawing on DAC work.

Reference groups, which will typically involve both Dutch civil servants and (paid) outsiders, including from the country concerned, are standard.

The number of impact evaluations is being increased from a low level, with a view to bringing them to at least half the programme. As explained above, the planning system facilitates this approach.

5) **Are findings disseminated in a way likely to secure impact?** [DAC Questionnaire 7.1, 7.3 and 8.4]

All IOB reports are submitted by the relevant Minister to Parliament with his/her policy response. This automatically gives some public visibility, particularly if a report is critical.

In most cases, there will be internal presentation of the findings of each evaluation, both at the weekly meeting held by the Director General with the Directors of the Development Cooperation Directorate (for development evaluations), and perhaps a brown bag lunch. Local presentations will also be held in recipient countries. However, IOB see it as the responsibility of the operating department to make most of the running. A recent exception was the Africa report, where IOB arranged a major conference at Rotterdam for 500 people, attended by both Ministers (Foreign Affairs and Development), to discuss the findings.

In its annual report to Parliament the ministry is obliged to include a section on the extent to which evaluation findings are implemented in policies. The Head of IOB does not himself publish an annual report.

FEZ consider that while the quality of IOB studies is good, their presentation leaves something to be desired (too dry and detailed), and that their high degree of independence works against ownership of their findings by operational departments.

6) **Are recommendations followed up systematically?** [DAC Questionnaire 7.2 and 8.3]

The Minister's accompanying report to Parliament follows of course discussion and advice from the operating departments being evaluated. The Minister's own view will be crucial to how seriously recommendations are taken. There is no very organised system for checking whether action is in fact taken. A proposal in 2008 by DEK to put in place a stronger follow-up system has yet to be taken

forward, not least because of a government-wide campaign to 'de-bureaucratise' the workings of government. The study now ongoing on the use of evaluations may reopen this discussion.

7) **How are evaluation products perceived by outside stakeholders?** [DAC Questionnaire 4.5, local stakeholders; 5.4 and 7.4 development agency staff; 8.1-8.2 other stakeholders, including Parliament and public]

As noted above, a feature of the Dutch system is the lively interest of some Parliamentarians, including those sceptical of aid, in the findings of evaluations. IOB reports appear to have public credibility, though that does not mean that they will get a high profile in many cases.

8) **What role does the evaluation department play in relation to evaluative material commissioned by other parts of the agency?** [Not covered in DAC Questionnaire]

IOB reviewed decentralised evaluation in a report published in 2004. This concluded that the quality of such evaluations was very variable. Of 25 studied in some depth, six rated good or very good on almost all points, eight were in the middle, and eleven rated only Fair or Poor. Many were basically arranged to inform management decisions whether to continue activities for a further phase, but then concentrated on how activities were progressing rather than on what impact the activities were having.

IOB provide a 'help-desk' function for decentralised evaluations, but do not aspire to a quality assurance role. There is a database of such evaluations, but it is not at all complete. In addition, senior Ministry staff feel that operational units and Embassies are probably commissioning an unnecessary number of evaluations, when lighter reviews would be more appropriate.

ANNEX 11: COMPARISON OF RESOURCES FOR EVALUATION: DFID AND COMPARATOR AGENCIES

	Key Statistics	DFID	EC	Irish Aid	IOB	NORAD	SADEV	SIDA/UTV
1	**Budget for studies, latest available year**	£4.3 m	€6.0m	€0.3m	€2.5m	NKr 25 m	SEK18m	SEK18m
2	**Bilateral ODA, national currency, 2008**	4111.2	6968	626.92	3696.72	17506.74	31593.91[17]	21584.45[18]
3	**Evaluation managers**	9	11	2	10	8	6	6
4	**No of central studies, typical year**	25	17	5	About 5	11	7	6
5	**Of which, joint studies led**	6	1 to 2	1	2	1	_	1
	Key Ratios							
6	**Studies Budget as proportion of bilateral ODA (1/2)**	0.11	0.09	0.05	0.08	0.14	0.06	0.08
7	**Average cost of studies (1/4)**	£172,000	€353,000	€60,000	€500,000	NKr2.27m	SEK2.57m	SEK3.0m
8	**No of studies per manager (4/3)**	2.8	1.5	2.5	0.5	1.4	1.2	1.0
9	**Bilateral ODA per manager (national currency) (2/3)**	£457m	€633m	€313m	€397m	NKr219m	SEK526m[19]	SEK390m

[17] Total, rather than bilateral ODA
[18] Bilateral ODA from Sida is somewhat smaller (some SEK 17bn), and this would be arguably a more suitable denominator in this case
[19] Total, rather than bilateral ODA

ANNEX 12: EVALUATION OF THE COMMISSION'S SUPPORT TO UNITED REPUBLIC OF TANZANIA: EXCERPT FROM 'FICHE CONTRADICTOIRE'

« FICHE CONTRADICTOIRE »

Evaluation of the Commission's support to United Republic of Tanzania

(*For details on the recommendations please refer to the main report)

Recommendations	Response of Commission services	Follow-up (one year later)
1. As focal sectors remain an obligation under the 10th EDF Programming Guidelines it is recommended to concentrate the funds in areas where EC's comparative advantages are well established: (1) **road network**, and (2) basic **education**	Broadly accepted – initial programming combines attention to transport facilitation (focusing on roads) and education (global sector approach in the light of ongoing reforms). It should however be noted that the programming process is still under way and that the future country strategy is defined jointly by the GoT and the European Commission together with the MS.	Transport will be a focal sector under 10th EDF, with sustainable financing of both the trunk and rural road networks a priority. However, the EC will withdraw from sector budget support to education: an increased overall package of general budget support and the comparative advantage of the EC in this sector will allow continued participation in sector dialogue.
2. In line with the Government's new strategy paper which is structured along poverty reduction outcomes, the evaluation recommends that the next EC strategy focuses on two focal outcomes: (1) **access to markets**, and (2) **equal and universal access to basic services**. Such a dual approach to concentration (2 sectors and 2 outcomes) is innovative.	Broadly accepted – a focus on pro-poor growth includes measures to improve market access and uptake of preferences (EPA and trade-related support, SPS, agriculture, transport), while access to basic services will be manifest in the attention given to quality and equity in education.	This recommendation is slightly overtaken by the fact that a joint (ie multi-donor) programming approach is taken in Tanzania. The Joint Programming Document agreed with government in 2006 gives the EC a leading role in pro-poor growth, to which its interventions in trade, agriculture and transport are especially closely linked. Sustainable growth is a prerequisite for the financing of reliable social provision. The joint programming document examines this relationship in greater detail.
3. The evaluation recommends that **governance** is considered as a prerequisite to all other outcomes, and is given the status of a top cross-cutting priority in the next country strategy.	Accepted. This is the approach followed in the programming of 10th EDF support to Tanzania	Governance issues including public financial management, anti-fraud, civil society participation and deepening democracy, play a critical role (cross-cutting and sectoral) in the EC's indicative programming as well as the joint programming document.
4. Address **future challenges**: technical barriers to trade potentially a priority sector in the future and vulnerability potentially a priority outcome in the future	Technical barriers to trade are to be addressed at both national and regional level, not just in funding but also in dialogue (SPS, trade facilitation, customs, ICT) Reducing vulnerability is only one of the 5 outcomes of poverty reduction. Given the focal sectors it is not sure that the future strategy will best be suited to reduce	TBT and other supply-side constraints receive focal sector attention in the 10th EDF programming. Follow-up of the Diagnostic Trade Integration Survey is a priority action for trade and private sector donors in Tanzania, including the Commission.

1

The UK Government is committed to ensuring that the UK's aid budget is used effectively to make a difference to the lives of the world's poorest people and that it represents value for money.

A new independent committee was set up to help the Department for International Development evaluate the impact of UK aid. The function of the **Independent Advisory Committee on Development Impact (IACDI)** is to strengthen and assure the independence of the evaluation function in DFID.

Although international development is already subject to extensive scrutiny, the Committee will challenge DFID'S thinking directly. The committee has important powers to:

- *determine which programmes and areas of UK development assistance will be evaluated*
- *check that international standards for evaluation are being applied*
- *monitor how recommendations are implemented*

IACDI works in a transparent way, publishing minutes of its meetings and consulting on the proposed future work plan of DFID evaluations on its website:
http://iacdi.independent.gov.uk/

The committee welcomes comments and suggestions on its work at all times and you can get in touch with the IACDI by email to:
mail@iacdi.independent.gov.uk