



Review of DFATD's Country Program Evaluation Approach

2013

Final Report

December 2013

Table of Contents

1. Introduction	1
1.1 Objective and Methodology	1
1.2 Current Approach to CPEs	2
2. Summary of Review Findings	4
2.1 Summary of Literature Review	4
2.2 Approaches of Other Donor Agencies	5
2.3 Review of DFATD’s CPE Reports	6
3. Issues with DFATD’s CPEs	8
3.1 Scope of CPEs: What Needs to be Assessed	8
3.2 CPE Methodology: How Should Assessments be Carried Out ?	12
3.3 Who Should do the Assessments ?	18
3.4 CPEs for What? (Raison D’Être)	19
4. Optional Approaches to Country Program Evaluation	22
4.1 Options: Seeking the right balance between accountability and learning, while addressing other issues	22
4. Conclusion	27
5. Annexes	28

Note to Reader

During the course of this review, the former Canadian International Development Agency (CIDA) was merged with the Department of Foreign Affairs and International Trade, creating the Department of Foreign Affairs, Trade and Development (DFATD). This change took effect on 26 June 2013. The former agency, CIDA, and its staff are referred to in this report as DFATD (implicitly the Development arm of the Department). However, references to documents that were published prior to the amalgamation retain the reference to CIDA.

Executive Summary

What is the Review?

DFATD had been conducting Country Program Evaluations (CPEs) for a number of years when, in 2008, it adopted Generic Terms of Reference and a Compendium of Tools to implement a standardized approach to these evaluations. After several years of use, DFATD decided to review the current approach to ensure it is in line with best practices and standards in the field of international program evaluation. This report assesses current processes and products for CPEs, reviews existing best practices in international development program evaluation, and presents options for discussion.

This report was prepared by DFATD Evaluation, based on data collection and analysis by DFATD Evaluation staff and Goss Gilroy Inc. consultants. It involved the review of relevant evaluation literature and DFATD documents, as well as interviews with senior managers; program staff; external members of the Evaluation Committee; consultants who conducted country program evaluations in the last three years; and representatives of bilateral development agencies in other countries (Australia, France, Norway, Sweden, United Kingdom, United States).

Review Findings

Interviewees in large proportion called for: maintenance of independently conducted country program evaluations; more emphasis on learning than on accountability; more timely evaluations to increase their usefulness to programs; greater focus in the evaluations on the program instead of projects; greater flexibility in the design and methodology; and improved usefulness of the analysis and recommendations. No group argued strongly for a standard methodology for CPEs.

Respondent groups generally agree that CPEs should continue to be undertaken by the corporate evaluation function and include a mix of internal (Development Evaluation Division) and external (consultant) resources to ensure independence and accountability. However, Evaluation Committee members tended to think the reports are not sufficiently critical of country programs and that the Development Evaluation Division (DED) does not demonstrate sufficient independence from the programming units. A major concern was expressed by all respondent groups (most strongly by the consultants) about the availability, in a timely fashion, of documents and high quality results data for the CPEs.

A number of the other aid agencies consulted do not undertake country program evaluations because they find them expensive, heavy in process and not sufficiently useful to decision-making. Many donors only cover components of country programs, others prioritize thematic or sector evaluations. Where done, country program evaluations are often led by programs managers. In the case of the European community, country reviews are done jointly by groups of member countries. Some agencies that currently practice only investment and thematic evaluations recognise that they lack the overall integrating narrative that the CPE provides. The literature

about country program evaluations, produced between 2004 and 2012, confirms that issues identified by respondents for this review are not new and are common to many aid agencies.

This review identifies issues (see section 3.0) that suggest the following adjustments to the CPE approach:

- more focus on depth rather than breadth of analysis;
- more thorough analysis of relevance and rationale related issues (beyond efficiency and achievement of results);
- better balance between learning and accountability;
- more timely evaluations to increase their usefulness to programs;
- greater focus in the evaluations on the program than on projects;
- flexibility and diversity in the design and methodology;
- the use of a simple rating system (if at all);
- options for level of effort depending on context (“heavy” vs “light” CPEs);
- improved usefulness of recommendations;
- improved measures for efficiency, economy and equity;
- renewed emphasis on neutrality and engagement of the evaluation division;
- clarity of roles and responsibilities between evaluation division and external consultants;
- improved availability, in a timely fashion, of documents and high quality data for data collection and analysis;
- more analysis of policy coherence across delivery channels at country level;
- increased use of contribution analysis .

Options

The report suggests options for consideration, intended to crystallise thinking on intended future use of country program evaluation:

- Option One: Accountability First (high accountability with some flexibility);
- Option Two: Variable Approach by Country (somewhat lower focus on accountability and more focus on learning for program improvement); and,
- Option Three: Decentralized and Integrated (priority on learning and utilization at program level).

As well there is analysis of different approaches to flexibility in issues to be addressed, the balance between evaluation of project and non-project activities, use of rating systems, the need or not for

a common approach and methodology, and the organizational arrangements for carrying out CPEs. (see section 4.0)

Next Steps

Following consideration of these issues and options by the Development Evaluation Committee, and further consultation within the Department, the Development Evaluation Division will undertake to develop next generation guidelines on approaches and methods for country program evaluation.

1 Introduction

The report reflects the results of a review of the methodology used by the development arm of the Department of Foreign Affairs, Trade and Development (DFATD) for conducting Country Program Evaluations (CPEs).

1.1 Objective and Methodology

DFATD had been conducting CPEs for a number of years when, in 2008, it adopted a Generic Terms of Reference (ToR) and a Compendium of Tools to implement a standardized approach to these evaluations. After several years of use, DFATD has decided to review the approach to ensure it is in line with best practices and standards in the field of international program evaluation, and recommend any improvements that may be appropriate.

This report assesses current processes and products for CPEs, reviews existing best practices in international development program evaluation, and presents options for consideration. Pending outcome of the consideration of issues and options, a subsequent step would be to articulate a revised approach, methodology and tools for country program evaluation.

This report was prepared by DFATD Development Evaluation Division, based on data collection and analysis by DFATD Evaluation staff and Goss Gilroy Inc. consultants. It involved:

- Review of relevant evaluation literature and DFATD documents (including CPE tools and synthesis and technical reports, Country Development Program Frameworks (CDPF), CDPF guidelines, documents related to results-based management (RBM) and risk management);
- Interviews with:¹
 1. DFATD senior managers (9) and program managers and staff (8);
 2. External members of the Department's Evaluation Committee (6); and
 3. Representatives of bilateral development agencies in other countries (Australia, France, Norway, Sweden, United Kingdom, United States) (6);²
- Discussion groups and some individual interviews were conducted with evaluation consultants who have experience in carrying out CPEs (14 participated).

For further information on the rationale and the methodology, see Annex A.

The report highlights 15 issues the review team considers appropriate to address to improve the current approach to CPEs. These relate to what needs to be assessed (CPE approach); how

¹ See list of interview respondents in Annex G.

² Some representatives of bilateral donors were not available for an interview but provided inputs by email.

evaluations should be carried out (CPE methodology); who should do the evaluations; and the overall purpose of CPEs. In section 3.0, the report provides a short description of each issue and suggests potential ways forward.

In section 4, the report presents three indicative options to help DFATD determine which direction it wants to emphasize in future CPEs. The options do not necessarily address all 15 issues. They do however provide a discussion framework for what approach the Department would like to emphasise. They range from a pre-eminent focus on accountability, to one that emphasises program improvement, and one that is run entirely by program managers. They include different approaches to flexibility, the balance between the evaluation of projects and non-project activities, the nature of a project rating system, the use of a common approach and methodology for all CPEs, and the organizational arrangements for carrying out CPEs.

1.2 Current Approach to CPEs

Overall Approach

The development evaluation function aims at “providing an evidence-based, neutral assessment of Department activities, to support accountability to Parliament and Canadians, as well as organizational learning and program improvement.”³ The Treasury Board’s 2009 “Policy on Evaluation”, and the Financial Administration Act, require that all direct program spending, including grants and contributions, be evaluated on a five-year cycle. This does not imply that DFATD must do CPEs, only that all expenditure must be evaluated.

There are currently three steps in carrying out a CPE:

- Planning phase – DFATD’s development evaluation manager develops a program profile and may conduct a scoping mission to the country. Based on this, the ToRs are finalized;
- Implementation phase – Consultants are hired to develop a detailed work plan and conduct a desk review. The consultants, and in some cases the DFATD evaluation manager, conduct a field data-gathering mission. After the mission and the data analysis, the consultants provide technical reports and in more recent practice an evidence binder;
- Reporting/communication phase – The consultants or the evaluation manager prepare a synthesis report and a one-page highlight sheet. Program managers prepare a management response. These documents are presented to the Evaluation Committee, which advises the Deputy Minister on quality, rigour, and any changes that might be required prior to approval. Following approval, the evaluation is provided to the Minister for information, and is subsequently published on line. One year after approval of the management response, and on an ongoing basis until completed, program staff report on implementation of the recommendations.

³ “DFATD Draft Evaluation Policy”, DFATD, 2012, p. 2

CPE Methodology

A set of CPE Generic ToRs were developed, tested and reviewed between 2008 and 2010. (See Annex B.) The CPE evaluation criteria were derived from the OECD Development Assistance Committee's (DAC) Quality Standards for Development Evaluations, the Paris Declaration principles and the Treasury Board Evaluation Policy. The criteria are: i) relevance, ii) effectiveness iii) sustainability, iv) cross-cutting issues, v) coherence, vi) efficiency, vii) management principles, and viii) performance management.

CPEs assess both a representative sample of projects, and the overall program, against the stated criteria. The project sample (usually less than 25 projects) should represent at least half of the Program's total budget, and its main sectors, thematic areas, and delivery modalities.

To assess the project portfolio by sector, and the overall program, CPEs use a rating system based on a five-point scale (see Box 1). The rating system can be used on a nominal basis (using satisfactory, unsatisfactory, etc.) and/or on a quantitative basis using the numeric intervals (e.g. 1.1, 1.5, 2). Annex B describes the CPE evaluation methodology in more detail and provides a summary of indicators used at the program and project level.

Box 1: Five-point Project Rating Scale

1. Highly satisfactory (4.1 to 5), the Program, project or initiative meets all the assessment indicators for the given criteria;
2. Satisfactory (3.1 to 4), the Program, project or initiative meets the main assessment indicators for the given criteria;
3. Moderately satisfactory (2.1 to 3), the Program, project or initiative meets some of the assessment indicators for the given criteria;
4. Unsatisfactory (1.1 to 2), the Program, project or initiative does not meet the main assessment indicators for the given criteria;
5. Highly unsatisfactory (0 to 1), the Program, project or initiative does not meet any of the assessment indicators for the given criteria.

2 Summary of Review Findings

This section summarizes the findings from the document review. In section 3, findings from interviews are presented for each of the issues identified.

2.1 Summary of Literature Review

Challenges related to CPEs are long-standing and known to donors.

Two key reports provide an overview of donors' experiences with country program evaluations. The first was conducted in 1999 by the Overseas Development Institute, in preparation for a high-level OECD-DAC conference in Vienna (Maxwell and Conway, 1999). The second, undertaken five years later, compared and contrasted the UNDP's experience with CPEs relative to that of other development agencies (UNDP, 2004). These two reports, with other (less extensive) analyses of donor CPEs during that period highlighted the following persistent challenges (Dahlgren, 2007; Flint et al, 2004; Cossée, 2001; Zurcher-Walther, 2006, World Bank, 2005; Sasaki, 2012):

- Attribution: difficulty of assigning the contribution of a specific aid agency to development results in a country;
- Usefulness: inward-looking and aimed at a limited range of stakeholders; trying to cover all the bases with limited time and space for detailed analysis; conclusions and recommendations too general;
- Design: overly long and broad evaluation designs, with too many projects included in the sample;
- Timing: limited connection between evaluations and the program planning process (relative to other monitoring exercises);
- Consistency: inconsistent interpretations and conclusions based on subjective assessments (rather than numerically measurable data);
- Overlap: lack of distinction between mid-cycle reviews (a monitoring activity) and country program evaluations;
- Ownership: limited sense of ownership and engagement by country offices and partner countries;
- Data quality: limited availability and quality of data to evaluate since some country programs are not developed with plans for how goals and objectives will be measured at the end of the cycle – a problem that stems from the lack of a results focus;
- Resources: high cost of undertaking regular CPEs;
- Clarity: if not clearly and diligently explained, the complex criteria and methods used in CPEs can lead to misunderstanding of findings and conclusions; and

- Ratings: inconsistencies between ratings and narratives.

This review illustrates the fact that donors have been dealing with challenges related to country program evaluations for many years.

2.2 Approaches of Other Donor Agencies

Many bilateral aid agencies have chosen to assess country programs otherwise.

Bilateral aid agencies from the Netherlands, UK, USA, Sweden, Norway, USA, Australia and France indicated that they have stopped doing CPEs because they are too expensive and complicated relative to their usefulness. Many donors only cover components of country programs, others prioritize thematic or sector evaluations. That said, some donors are again considering doing country program evaluations as a way of integrating their now fragmented project and thematic evaluation effort.

The Netherlands has only conducted one CPE in the last fourteen years. Denmark covers specific sectors or themes for a specific country (as opposed to full country program evaluations), often covered through joint evaluation with other donors and/or partner country. Norway only does CPEs on an as needed basis, based on specific program that require clarifications or explanations. France' aid agency (AFD) does few CPEs and only jointly with the French Foreign Affairs and Finance departments. They cover a wide variety of issues often beyond development. Occasionally, AFD would do joint CPEs with the European Commission. Australia's AusAID is currently considering the use of CPEs in a more systematic mode. It seeks a more comprehensive consideration of outcomes over a period of assistance that is long enough to provide concrete evidence of development results.

The United Kingdom's DFID central evaluation office no longer carries out CPEs. DFID's country offices are now expected to do their own evaluation work, including overall CPEs. The Independent Commission for Aid Impact, established in 2011, can do CPEs but usually does assessments more akin to country program performance audits. USAID has done some CPEs in the past. The Agency now prefers to look at specific program components across many country programs, or look at one given sector across many country programs. For accountability purposes, USAID does "management assessments," which focus on resource management. For learning purposes, the Agency focuses on program components or on sectors/themes. USAID and DFID's evaluation services provide support to country offices, as these now lead their own CPEs.

It is worth noting that some agencies (namely USAID, DFID and UNFPA) have approached the question of roles and responsibilities differently. They made CPEs the responsibility of their decentralized country offices (and/or operational desks at HQ) but kept the evaluation quality assurance and support role with their respective evaluation services. A decentralized CPE structure makes the organizational body responsible for developing and implementing programs also accountable for effective, timely and high quality evaluation. This is only practical if there is an effective system for quality assurance and oversight of CPEs.

Regional development banks have indicated that, given their country-based membership and their governance system, it makes sense for them to focus on country strategy/program evaluations. Among UN organisations, the practice of CPEs seems to have fallen from favour at UNICEF. A review of UNICEF evaluations carried out in 2012 found that only one country program evaluation had been carried out in the previous three years.⁴ In contrast, at UNFPA, the practice of CPEs has recently been revitalized with a new evaluation policy, a re-arrangement of the lines of authority for country program evaluations and a new set of guidelines.⁵ CPEs at UNFPA are a compulsory part of the program planning cycle and must be completed before a new program can be approved. UNDP continues to invest in regular CPEs although a review of their evaluation reports conducted in 2012 noted a decline in frequency over the prior four year period.

A key concern for aid agencies remains the amount of resources and time they should spend on CPEs. More details on the approaches of other donors can be found in Annex C.

2.3 Review of DFATD's CPE Reports

The review of CPE reports confirms some of the issues interviewees have identified (see section 3).

Since 2008 DFATD has completed, or is currently undertaking, CPEs for all its countries of focus (see Table 1, Annex F). In addition, it has conducted CPEs for two country programs that are not countries of focus. In total, it has conducted seventeen CPEs since 2008, with evaluations for nine more countries in progress.⁶ A detailed review of the reports from ten recent CPEs⁷ was conducted and the findings that provide key background information for this review were:

The key characteristics of the current approach to CPEs are:

- CPEs are centralized as a responsibility of the Development Evaluation Division (DED) and are paid from the operating budget ;
- CPE objectives and evaluation criteria are standard and are based on the generic Terms of Reference;
- CPEs are designed using a common methodology which encompasses common rules to be used in sampling projects in the country program;
- In assessing the sampled projects, CPE teams are relying on existing project evaluations (secondary data), even though it is recognized that these are often lacking or weak (too few or poor quality project evaluations; incomplete or missing documents, outdated or contradictory

⁴ "Review of UNICEF's Development Effectiveness 2009 - 2011: Final Report", OECD-DAC, 2013

⁵ "Handbook: How to Design and Conduct a Country Program Evaluation at UNFPA", UNFPA, March, 2012.

⁶ Ethiopia, Vietnam, Honduras, Inter-American Regional, Senegal, Ukraine, Caribbean Regional, Colombia, Mali, Peru and Bolivia. All of which are complete, except Bolivia.

⁷ Bolivia, Caribbean Regional, Colombia, Ethiopia, Honduras, Mali, Peru, Senegal, Ukraine and Vietnam.

statistics);

- For a few years, in assessing sampled projects, CPE teams applied a common rating scale to each evaluation criterion, and averaged the numeric results across criteria for program and projects (this practice has been abandoned in 2011, given the poor value added of such averages to analysis);
- The current approach emphasizes mainly project assessments at the expense of non-project activities, including strategic issues related to the programming sectors, and investments;
- CPEs are conducted between one and three years after the end of a program planned period, as per the Country Development Programming Framework;
- All CPEs cover bilateral, multilateral, and partnership delivery channels;
- Evaluations address the sectors of concentration in a given country, and gather information at both the project and program levels. They do not include sections on policy dialogue and capacity building unless there are components in projects and/or the program related to these;
- For nine of the CPEs reviewed, there were problems with the availability of, and timely access to, data and documents. There were also limitations in accessing national data in country (Caribbean, Colombia, Mali);
- CPEs make extensive reference to program logic models, Country Development Programming Frameworks and Performance Management Frameworks. However, issues frequently identified with these corporate documents include the fact that they are not updated to keep in line with changes in the country context or internal policy or priorities, and baseline or monitoring data is not collected on indicators;
- Different approaches are used to presenting conclusions and recommendations. The format is not systematic across all CPEs. Some reports do not have conclusions, only findings that lead to recommendations or suggestions. Also, recommendations are sometimes called “suggestions” or “next steps” – sometimes at sectoral level, sometimes linked to each evaluation criterion.

3 Issues with DFATD's CPEs

This section draws out 15 key issues identified in the data collection and discusses their implications for the approach, methodology, and report format of CPEs. The issues relate to what needs to be assessed (CPEs scope in Section 3.1); how the assessments should be carried out (CPE methodology in Section 3.2); who should do the assessments (Section 3.3); and the purpose of CPEs (CPEs raison d'être in section 3.4). These 15 issues are not present within all CPEs. Rather, they are a summary of issues that emerged across CPEs.

3.1 Scope of CPEs: What needs to be assessed.

A number of issues were raised in the interviews and confirmed in the document review with respect to the scope of CPEs: what it currently is and what it should be. What should be covered by evaluations, should they be undertaken in all program countries, should both project and non-project activities, and all delivery channels, be covered?

Issue 1: Treatment of Relevance and Rationale

CPEs could address essential programming matters.

There is partial disagreement between views of management and program staff and external members of the Evaluation Committee with respect to the way that program relevance and rationale is treated. Senior managers (Assistant Deputy Ministers, Regional Directors General, Directors General) and program Directors and managers tended to think that it was not appropriate or, in fact, useful for the evaluations to address the rationale⁸ of a program, or the choice of projects, in a given country. They alluded to ministerial prerogative, and the fact that country and Canadian contexts change over time. On the other hand, members of the Evaluation Committee felt strongly that, in order to be truly useful to the Department, the CPEs should be more strategic, focussing on such questions as the choice of sectors, the choice of projects, and how components of the program work together.

Observation:

CPEs could become more useful for programming purposes if they addressed strategic programming issues, including the choice of sectors and projects, as well as sector related best practices. It is part of the evaluation function to provide analysis for decision-making, including such basic questions as: are we doing the right things? Are we doing them right? Are they having the desired influence or impact?

Issue 2: Use of CPEs in all Country Programs

Not all country programs need to be assessed the same way.

⁸ Beyond the current assessment of the alignment of the program and projects with Government and DFATD-Development priorities.

Most interview respondents thought that CPEs would be appropriate in all country program contexts – low- and middle-income countries and fragile states. However, there was recognition that adaptation would be required in some country contexts. Interviewees suggested that not every country program needs to be assessed in the same way. Countries of modest presence, for instance, could be grouped by region into a series of regional evaluations. Fragile states could be combined into one or several thematic evaluations (e.g. focus on displaced persons, programming in state-in-conflict), so that conclusions contribute to improving future programming. Countries of focus could also be merged to explore common programming experiences by sector, or to explore the use of common delivery mechanisms. Another possibility would be to consider if, while some key country programs might be worth a detailed evaluation, others might require only a lighter form of CPE, which would focus solely on program-level issues, and discard the assessment of project-level investments. The overall evaluability assessment of a country program (in addition to purpose and scope), would also inform whether a ‘lighter’ or ‘heavier’ CPE is undertaken.

The Inter-American Development Bank has an interesting typology of country assistance evaluations, each type addressing specific needs: country studies (broad assessment of the political, economic and social situation of a country); country strategy evaluation (an assessment of the relevance and effectiveness of the Bank’s strategy in a given country); country program evaluation (an assessment of the relevance, effectiveness and efficiency of country-level operations); country assistance evaluation (an assessment that includes program sustainability and contribution).

Observation:

Generally, the evaluation could be ‘lighter’ or ‘heavier’ depending on the structure of the country program, the evaluation issues being addressed, the amount of programming covered, the time-period under review, the forms of the donor-partner relationship being examined, and the overall evaluability assessment of the country program in question. While ‘heavier’ program evaluations produce more conclusive findings, and allow more comparative review of country program performance, ‘lighter’ evaluations can produce policy-relevant recommendations rapidly, which is particularly useful when the aid program is set within changing circumstances. Light versions of program evaluations typically cover selected program components, a limited selection of evaluation criteria (often: relevance, efficiency, coherence, effectiveness), and cover a shorter period of programming. Also, a light evaluation can be designed to focus on program-level matters only (policy dialogue and aid effectiveness practices, for instance) and not include the assessment of projects.

Issue 3: Coverage of Project Investments and Non-project Activities

A better balance in coverage would enhance CPE reports’ analysis.

It is widely recognised that a program is more than the sum of its projects. Interviewees generally agree that CPEs should include non-project activities. CPEs often pay limited attention to these, in comparison to the time and resources spent on assessing projects. In fact, most aid agencies lack well-developed results frameworks for non-project activities. Donors do not systematically measure the impact of policy dialogue (for instance) and do not document their policy dialogue

activities, nor do they have specific guidelines or models. In response to this trend, DFATD has begun to develop tools to better assess the contribution of policy dialogue activities within country programs (See Annex D for a further discussion on assessment of policy dialogue activities, including references.)

CPEs currently focus on measuring results of project investments. It is equally important to understand how those results have been achieved. This would require including in the theory of change how non-project activities were expected to have an impact on the overall results.

Observation:

Better coverage of non-project activities requires refocusing the current generic terms of reference for CPEs and developing corporate indicators related to policy dialogue and capacity building. Also, incentives for staff to include non-project activities in program/project planning and monitoring may be required to ensure CPEs are subsequently able to assess these activities.

Issue 4: Coverage of All Delivery Channels

Continue to include direct support to multilateral and non-government partners in CPEs, but focus more on alignment and policy coherence.

An important question regarding the scope of CPEs concerns whether or not they should include the evaluation of direct (non-country specific) support to multilateral agencies and non-government partners. Interviews produced a mix of opinions. Some felt that a genuine evaluation of support to a given country should include all forms of support, whether direct or indirect, including the programs of multilateral agencies provided with core or program specific financial support. The same logic would apply to activities in a given country by a Canadian partner organization or its local affiliate, if that partner was receiving support from DFATD.

Other interviewees argued that CPEs should focus only on bilateral activities. This would involve evaluating only activities planned and funded under Canada's bilateral program of government-to-government assistance as described in the Country Development Programming Framework and managed within DFATD. This would include activities of multilateral agencies and Canadian and local NGOs, provided they were funded by the bilateral program (multi-bi funding, in the case of multilateral agencies).

If DFATD evaluates the programs implemented by a given multilateral agency (which are not directly funded by the department under a multilateral-bilateral arrangement), it will, to some degree, substitute its own accountability or that of the agency involved. The primary responsibility for evaluating multilateral programs at country level rests with the agencies themselves and with their government partners. It may also put the results of DFATD's evaluation work in conflict with the results reported by the agency's own monitoring and evaluation function. On the other hand, strictly limiting the CPE to the review of projects and programs funded directly by the bilateral program (including multi-bi projects) in a given country would run the risk of ignoring some very important aspects that may affect the program's effectiveness.

In virtually every country where DFATD has a bilateral program, its staff interact with, and monitor the work of, multilateral agencies and NGOs. In most countries DFATD staff take part in consultative mechanisms to ensure that all parties contribute to a common development agenda in a coherent way. At relatively low cost, CPEs could examine how DFATD's programs align or conflict with the work of other key development actors: bilateral, multilateral and non-government. It could also examine the extent to which DFATD's advocacy and policy support activities are coordinated with those of other agencies, particularly influential multilateral agencies.

Observation:

Comparisons of the effectiveness of delivery channels are challenging, given differences in approaches and assessment of results. However, the CPE could examine, as part of its assessment of non-project activities, how DFATD's programs and advocacy/policy dialogue align or conflict with the work of other key development actors, including multilateral and non-government organisations. Eventually, DFATD may wish to address the issue of whether and how the CPE focus could integrate an assessment of the country-specific objectives for diplomacy and trade.

Issue 5: Value for Money

CPEs could make good use of improved measures for efficiency, economy and equity.

Value for money (VfM) has become more prominent on the development agenda because aid agencies are increasingly expected to understand and demonstrate VfM to taxpayers. Strong evidence is needed to demonstrate that aid is valid and well-managed. The issue has been raised several times during this review process and is therefore worth exploring.

The OECD defines VfM as the optimum combination of cost and quality to meet the user's requirements. It is usually assessed using the criteria of economy, efficiency and effectiveness. However, VfM in development evaluation is now being more frequently referred to as the best balance between economy, efficiency, effectiveness and equity.⁹

Annex E provides further information about Value for Money (VfM), in addition to different approaches to value for money. Donors' experience shows that the integration of VfM in corporate operations requires good planning, pragmatism and flexibility. The absence of VfM analysis is no better than rigidly applied VfM. However, according to donors' experiences, a clear corporate agreement must be achieved about value for money for whom, of what and by when, to avoid misinterpretations about the purpose of VfM, resistance about the use of tools, and the design of complex and expensive monitoring systems.

Treasury Board defines value for money (VfM) as the extent to which programs demonstrate *relevance* (continued need for the program, alignment with government priorities, alignment with federal roles and responsibilities), and *performance* (achievement of expected outcomes,

⁹ Jackson, Penny, Value for money and international development: Deconstructing myths to promote a more constructive discussion, OECD Development Co-operation Directorate, May 2012.

demonstration of efficiency and economy). In its CPEs, DFATD systematically uses all of these criteria. However, the review of CPE reports shows that the measurement of efficiency and economy, although challenging, could be strengthened.

Observation:

DFATD is currently developing corporate indicators to improve its ability to measure efficiency and economy and be more in line with Treasury Board's definition of Value for Money. Given DFATD's development mandate, the "equity" criterion could also be explored.

3.2 CPE Methodology: How should assessments be carried out?

A number of issues were raised in the data collection about components of the CPE approach and methodology. These range from the fundamental question of attribution to questions of qualitative vs. quantitative methodologies. Each is explored in this section.

Issue 6: Results and the Problem of Attribution

Contribution analysis can be an appropriate approach to country program evaluations.

While the current CPE methodology focuses directly on the evaluation of program components and addresses the core Development Assistance Committee evaluation criteria, it does not directly answer questions related to attribution. While assessing what results were achieved, the extent to which results can be attributed to DFATD-supported projects and programs is not usually covered. The problem of attributing results has been given increasing prominence in development evaluation over the past decade.

Contribution analysis is a useful tool (Treasury Board promotes its use) in situations where an agency's support is just one part in a complex network of international and national actors influencing developmental outputs and outcomes.¹⁰ This would seem to fit DFATD's needs in many countries where it provides support to sectors and fields of activity where other, more financially significant, donors are almost always present. Rarely would DFATD have a given field of development to itself as the source of external policy, material or financial support.

As the literature makes clear, contribution analysis can be applied at the activity, project, or program level.¹¹ Theoretically, a CPE could conduct separate contribution analyses for each of the thematic areas of focus in a given country program. It could also include a contribution analysis for the program as a whole. This pre-supposes, however, that the program as a whole can be

¹⁰ "Contribution Analysis: An Approach to Measuring Cause and Effect". John Mayne, Institutional Learning and Change (ILAC) Brief 16, May 2008. Consultative Group on International Agricultural Research (CGIAR).

¹¹ "Contribution Analysis: An approach to exploring cause and effect". The Institutional Learning and Change (ILAC) Initiative: ILAC Brief Number 16, May 2008. Published by the Consultative Group on International Agricultural Research (CGIAR)

described in a reasonable, credible theory of change. A number of DFATD interviewees have pointed out that country programs evolve over time and react to many different and changing policies and priorities. It might be very difficult to discern a reasonably cohesive theory of change as the basis for any given DFATD country program.

Observation:

A properly applied contribution analysis would help to address the problem of attribution as part of a well-designed CPE, assuming the evaluation has access to adequate information on results. However, effective contribution analysis pre-supposes that the intervention under review can be described by a cohesive theory of change and has made a reasonable effort to measure and report on program results.

Issue 7: Whether or not a common methodology is needed

The need for comparisons does not preclude diversity of methodologies.

DFATD has designed a single approach to its CPEs to ensure greater comparability between programs (as described in section 1.2 and in annex C). No group during interviews argued strongly for a standard methodology for CPEs, and all felt that there was a need for more flexibility to ensure the usefulness of the evaluations. Few comparisons across programs have been done¹². Some challenge the value added of comparing performance between country programs, given the wide diversity of programs and country contexts. However, many think that comparisons of specific issues between programs (management of gender equality operations for instance) could be relevant. Such comparisons do not preclude diversity of methodologies.

If more flexibility is to be encouraged in the issues addressed by CPEs, then there is less scope for a common methodology. Among factors that could affect the methodological design of a given evaluation are: the type of country; program context; sectors; size of program; existence of particular delivery mechanisms such as budget support; the availability and quality of information; and management's interest in a particular issue.

Observation:

For most donors, a standard methodology applied to all country program evaluations is not considered good practice. Corporate evaluation services have diversified the design of their evaluations to address a narrower set of issues, within a wider variety of programs, both in size and in nature. Given the need for flexibility, aid agencies prefer to emphasize common standards of quality that ensure consistency among country program evaluations. This balance between common standards and flexibility ensures rigour, while not imposing rigidity. The Development Assistance Committee and Treasury Board do endorse the use of different methodologies to assess their prescribed evaluation criteria.

¹² A key exception here is the Grants and Contributions Review (2006, 2011), which aggregated investments across CIDA.

Issue 8: Rating System

Keep rating scales easy to use for evaluators, easy to understand for users, and relevant to the purpose of the evaluation.

Interview respondents agreed that methodologies currently used for collecting primary and secondary data (document and literature reviews, interviews, focus/discussion groups etc.) were all considered to be appropriate. It was the project rating system that caused the most concerns.

Although there were some who felt that the project rating scale brings more structure to the CPEs, generally senior managers and Evaluation Committee members were sceptical of the rating system because of the lack of a clear link to supporting evidence for the ratings, resulting in a spurious sense of rigour in the evaluation. Evaluation managers and consultants also expressed concern about the lack of definition in the project rating system, leading to considerable subjectivity in the ratings and contributing to a sense that, although ratings are given, the information in current CPEs being used for accountability purposes may not be very strong.

Canada is one of the few donors currently applying a rating scale system to country program evaluations.¹³ Others have taken the view that a focus on ratings can constrain flexibility to investigate context specific issues, and limit richness in data collection, analysis and narrative. On the other hand, where no clear conclusions emerge regarding a given sector or program, a rating does provide a tangible outcome and indication of accountability for results.

Among those donors who use a rating scale, the 4-category traffic light system is a noteworthy model (see Box 2). It is used by the World Bank's performance Annual Report, the UK Independent Commission for Aid Impact, and is similar to DFATD's current Annual Country Reviews. No numerical rating is assigned to the nominal categories, only colors are used to indicate performance. This option would reduce (to some extent) the richness of data, but the smaller number of categories could also improve consistency and comparability, in addition to being easy to understand by intended users of evaluation reports. Removing the mid-point category forces clear response and results that facilitate decision-making.

Rating scales do not need to be used in all CPEs.

Criteria could be developed that determine when rating scales will be used in CPEs, i.e. in heavy evaluations oriented toward accountability, rather than lighter evaluations where the focus is on learning. Regardless of the type of rating scale DFATD chooses to retain in its CPE, the principles of consistency, comparability, clarity, learning-

Box 2: a few donors, including DFID, use a rating scale centered on a traffic light color system.



¹³ Sasaki (2012).

oriented, user-friendly, flexibility, and emphasis on analysis should guide its use.

Observation:

The value of a rating scale system should be weighed against accountability and learning needs. If rating scales are to be retained, detailed guidelines for their implementation should be developed. These should include a clearer definition of the criteria for ratings, both at project and program level, and prescribe a common process for developing the ratings, in order to ensure consistency in application across evaluation teams and countries. Clearer definitions for criteria would help reduce inter-evaluator variability in scoring.¹⁴ In addition, it would provide the basis for a stronger link between evidence and findings.

Issue 9: Availability of Data and Documents

An issue with serious consequences for the quality of CPEs.

A major concern identified by those who were involved in conducting CPEs (evaluation staff and consultants) was the challenge of obtaining data and documents relevant to the country program. The methodology relies heavily on secondary data sources, particularly for the project ratings, with field missions used to confirm findings identified in the document review. However, there are multiple problems with this secondary data:

- There is the perception that there is potential bias in the project reports developed by program staff;
- There is limited results-based management data available at the project level. The lack of a results-based framework and/or information and data was noted specifically in five CPE reports (Caribbean, Colombia, Ethiopia, Honduras, Mali);
- Not necessarily a significant sample of projects have been subject to external monitoring or evaluation. Gaps in project evaluations were specifically noted in two CPE reports (Peru, Senegal);
- Even for projects that have been evaluated, the evaluation team is not always able to find the evaluation reports, or find them in a timely fashion. There is considerable variation in the number of documents, and the way they are stored, in the document database at DFATD. Concerns with DFATD's information systems were mentioned in six CPE reports (Caribbean, Colombia, Bolivia, Peru, Senegal, Ukraine);
- Project evaluations, even when available, are not always considered as useful because they focus on process rather than results;
- The search for these documents cannot be done by outside consultants, so it places a burden

¹⁴ It would be important that these definitions take into account the possibility of very different conditions (particularly in terms of the evolution of the program) in each country.

on program staff to find the documents which often results in delays in the first part of the evaluation - data collection; and

- Gaps in financial data from DFATD were noted in two CPE reports (Colombia, Mali).

These challenges with the availability of secondary data represent a major limitation to an evaluation methodology. This limitation becomes even more significant if the Department moves to include contribution analysis in its CPEs.

Observation:

Problems of access, collection, organization and archiving of documents limit the evidence on program and project performance. CPEs cannot be credible without ready access to good quality data. There should be renewed consideration of appropriate coverage of project evaluations during the program cycle (to be assessed for quality by the evaluation division), as well as ongoing data collection by the country program of information on results at the output and outcome level.

Issue 10: Quality of Recommendations

There is no consensus on the purpose of recommendations

Program managers and staff underlined that recommendations are not always targeted at things that are doable within the existing context. Proposed recommendations do not always recognize changes that have occurred in the program since the evaluation was conducted. External Evaluation Committee members expressed concerns about the potential lack of independence of consultants in making recommendations – linked to the lack of critical recommendations. Recommendations tend to be instrumental and amount to tinkering. Often they are not aligned with findings which were more critical. In addition, confusion remains about how to deal with corporate-level recommendations. Some question how there can be recommendations at corporate level be made on basis of one CPE, but others see common issues (e.g. decentralization, delegation of authority) worth exploring, because they have direct impact on program performance.

Observation:

Recommendations should serve both operational management and higher level strategic decision making. They currently do not do the latter to a reasonable extent, being mostly technical in nature. Fundamental and/or critical issues should percolate up from evaluations, including issues common to several programs.

Issue 11: Evaluation Criteria

Definitions of evaluation criteria bring confusion, criteria are misused and there are too many

There is overlap between criteria (relevance and coherence, efficiency and performance management, effectiveness and sustainability) often resulting in some evidence being reviewed and included in ratings twice. The analysis between criteria in CPE reports is often similar and therefore repetitive, in particular in relation to cross cutting themes. The overlap between project and program criteria is important and is evident in the ratings. Many consultants underlined that they are not given sufficient time to properly cover every single criterion, and stressed the need for greater flexibility in definition of criteria for better country specific analysis. A rigid use of evaluation criteria produces mechanical reports.

Observation:

Current CPEs tend to be demanding in breadth at the expense of depth of analysis. A reduced number and better defined criteria would give evaluators the opportunity to emphasize depth and quality of analysis. The proposed approach, which can be refined as needed, would reduce the number of evaluation criteria to five, organised in two categories: 1) Development results (relevance, effectiveness, sustainability); and 2) Management factors (efficiency and economy). These evaluation criteria will be informed by Paris principles for aid effectiveness. Cross cutting themes of gender and environment will also be addressed. Finally, 'coherence' may need to be considered in the context of the new DFATD.¹⁵

Issue 12: Timeframe for CPEs

Timeliness of CPEs, a corporate responsibility

The timeframe for completing CPEs was a major issue raised in the interviews. The average timeframes are:

- Thirteen months for the completion of work plan, data gathering, technical and synthesis reports. This does not include the time required for the inception phase, including contracting and finalizing the Terms of Reference; and
- An average additional 8.5 months for management response, committee review, approval, and publication of the CPE report.¹⁶

Evaluation staff and consultants identified a few reasons for this:

- The length of time for contracting the consultants;
- The time to access DFATD documents and data; and
- The length of time required for all stakeholders to review and comment on the draft technical and synthesis reports.

¹⁵ Although the term [coherence] has been increasingly used in development cooperation debates and statements of differing legal standing, there is no accepted definition and operationalisation of coherence in development evaluation methodology. Keijzer, N. and J. Oppewal (2012)

¹⁶ This average is based only on the nine CPEs completed to date.

Observation:

Current procedures need to be assessed to determine how they can be streamlined to ensure the CPEs are conducted in shorter timeframes to meet program planning timelines, while still ensuring adequate quality assurance. In addition, it should be noted that the Audit Division is now required to produce Country Program Audits on a seven year cycle. In developing new instruments, the Evaluation Division will need to take into account situations where an audit and an evaluation could potentially be done at the same time for a given country program, and examine how they can be coordinated.

3.3 Who should do the assessments?

Issue 13: Neutrality and Credibility of CPEs

Some CPEs may be too accommodating.

Interviews, particularly those with external members of the Evaluation Committee, indicate there is significant concern that the findings and conclusions of CPEs may be influenced by DFATD program staff. Evaluation Committee members felt that the Development Evaluation Division staff responsible for writing CPE reports (and presenting the results of CPEs to the Committee) are too reluctant to challenge the opinions of program staff and to defend their findings and conclusions. They pointed to a number of potential causes:

- The close relationship which develops between the evaluators and the program during a CPE, with the evaluation team reliant on the program for access to information and data and for logistics and other support;
- The vigorous and sustained pushback from senior program officers both before and during presentation of CPE results to the Evaluation Committee;
- A relatively weak underlying foundation of evidence in some CPEs, particularly in the form of weak or absent project evaluation and monitoring data, which undercuts the evaluators' ability and willingness to make clear findings and conclusions and to defend them.

Respondents felt that these factors contributed to their sense that CPEs tended to be overly positive and to avoid making substantive critical observations on program effectiveness, even when the underlying project and program information would support such observations.

Observation:

Findings underlined the importance of the evaluation function's neutrality, but also of its engagement with programs for the purpose of understanding context and identifying key issues. In addition, solid evidence "binders" are key to support CPEs findings and conclusions and challenge pushbacks from programs.

Issue 14: Roles and Responsibilities for CPEs

Roles between internal and external evaluators to be reviewed.

Almost all the interview respondent groups felt that the current set of basic responsibilities for CPEs at DFATD was appropriate. Specifically, they felt that the Development Evaluation Division (DED) at headquarters should be responsible for commissioning and carrying out CPEs with the appropriate program division acting as a key internal stakeholder. Some interviewees noted that this arrangement was especially appropriate given the importance of CPEs as an instrument of program accountability.

There remains the question of how to staff and field the evaluation team. Most importantly, for those interviewed, what should be the split in roles and responsibilities among DFATD staff and contracted evaluation consultants?

One way to think of this question is whether DED staff should be seen as the commissioners and managers of CPEs, or should they take an active role in all aspects of the CPE, including primary data collection in the field and analysis of primary and secondary evaluation information during preparation of draft and final reports. Interviews with evaluation consultants indicate that, from their perspective, having DFATD evaluation staff take part in direct data collection during the main evaluation mission causes a number of practical problems in evaluation team management (what is the role of the external team leader vis-à-vis DFATD staff, for example). Overall, maintaining neutrality on the part of internal evaluators, but also ensuring an adequately nuanced understanding of the corporate environment for consultants, are both essential for delivering credible and useful evaluations.

Observation:

Staff from the Development Evaluation Division (DED) should retain full responsibility for the preparation of Terms of Reference and the development and management of the Request for Proposal and the entire proposal process. During the Inception Mission, the optimum solution is shared involvement of DED and consultant staff, involving a joint inception mission by the DED evaluation manager and the consultant team leader. For data collection, a reduced level of involvement of DED staff is preferable, with DED staff not usually taking part in interviews and focus groups. See Table 3, Annex F.

3.4 CPEs for What?

Issue 15: Accountability vs Learning

The right balance between accountability and learning needs to be found.

Overwhelmingly DFATD staff and Evaluation Committee members felt DFATD should continue to do CPEs. Organising evaluation of geographic programs in this way corresponds with the management structure currently employed for program delivery. Most respondents agreed that having CPEs as the centrepiece of DFATD's geographic program evaluations does not preclude it from conducting evaluations of geographic programs with a focus on a sector, theme or programming approach when the need to address specific issues arises.

In order to align with the Treasury Board Evaluation Policy and the Development Assistance Committee criteria, DFATD addresses a number of specific issues in its evaluations. Under the current CPE methodology this means that a lot of the CPE “space” (in terms of budget and time) is taken up with pre-determined issues as stated in the terms of reference. CPEs seem driven almost exclusively to meeting the accountability requirements, measuring the “ends” reflected in the projects. The focus of evaluations on issues that are not necessarily relevant to program managers has made it difficult to link the CPE results to program improvements (learning).

Accountability implies two things: being responsible for ensuring that something happens and being able to show that the responsibility has been fulfilled.¹⁷ Accountability, as a purpose for evaluation, focuses the evaluation on what was intended to be achieved and the measurement of what was, in fact, achieved. On the other hand, the focus of learning evaluations is on issues of relevance to future programming that can help inform decision-making by program managers.

Senior program managers are interested in both accountability and learning, as they focus on demonstrating achievements in the areas for which they were responsible and for ensuring that program managers are learning (and applying the lessons learned) from evaluations. Many interviewees underlined the need for more flexibility to address issues of relevance to program managers, such as innovative programming and replicability of projects.

Observation:

CPEs are one among many corporate mechanisms that address accountability. Other types of evaluations and audits do so as well, in addition to program annual reports. Treasury Board confirmed to the Review Team that greater focus on learning is acceptable as long as there is some coverage of the two core issues of relevance and performance.¹⁸ The CPEs potential shift to incorporate more aspects of learning has implications for the way in which the CPEs would be planned and conducted. This includes the timeliness of evaluations, their flexibility to include specific issues of relevance to programs, more time spent on country and program context analysis, and more time spent on research and analysis of programming alternatives or good practices. See Table 2, Annex F. This last issue prompted the design of the three options described in section 4.

The table below summarizes the main characteristics of the current approach to country program evaluations:

Strengths	Weaknesses
-----------	------------

¹⁷ Drawn from “Accountability in International Development Aid”, Leif Wenar.

<http://www.wenar.info/Accountability%20in%20International%20Development.pdf> Accessed September 2013

¹⁸ Directive on the Evaluation Function, Treasury Board Secretariat, Annex A “Core Issues to be Addressed in Evaluations”, 2009

<ul style="list-style-type: none"> • Direct emphasis on accountability • Use of same (or a very similar) template to measure the results and effectiveness of every country program • Can incorporate a common methodology • Calibration of findings against a defined set of common criteria • Numerical score for each criterion for each project • Provides more structure to the evaluation than existed before the project rating system was introduced • Provides a structured, quantified assessment of project level results which can support Value for Money (VfM) assessment 	<ul style="list-style-type: none"> • Lack of flexibility (mainly due to the time and effort required to apply the methodology systematically to each sampled project) • Little attention given to the non-project activities in country programs • Limited ability to incorporate issues and criteria which respond to the more immediate needs of the program • Challenge of applying the project rating system consistently in all contexts • Difficult (given resource requirements for assessing project level results) to use the model to address program innovations or to adopt a risk-based approach • Less appropriate for addressing issues of scaling-up or replicating program innovations
--	---

4 Optional Approaches to Country Program Evaluation

Issues to be addressed relate to what needs to be assessed; how the assessments should be carried out; who should do the assessments; and the purpose of CPEs. To help determine which direction to take in relation to CPEs, the Review Team has developed three options. These range from a focus on meeting senior management's needs for external accountability to a focus on meeting program improvement needs of program managers. They include different approaches to flexibility, the balance between the evaluation of projects and non-project activities, the nature of a project rating system, the use of a common approach and methodology for all CPEs, and the organizational arrangements for carrying out CPEs. For ease of reference, each option has been assigned a title:

- Option One: Accountability First (high accountability with some flexibility);
- Option Two: Variable Approach by Country (somewhat lower focus on accountability and more focus on learning); and,
- Option Three: Decentralized and integrated (priority on learning at program level).

4.1 Options: Seeking the right balance between accountability and learning, while addressing other issues

Option One: Accountability First

Option One maintains accountability as the primary purpose of CPEs. However, taking into consideration the issues outlined in section 3.0, it adds certain elements that would have the effect of increasing flexibility and allowing some improvement in the learning/utilization aspect of CPEs. It would retain the first four characteristics of the current model: centralized responsibility for CPEs at Development Evaluation Division (DED); common evaluation criteria; common methodology; and, a strong reliance on existing project evaluation reporting where it is available, supplemented by some primary and secondary data collection in the field.

Where Option One differs from the current approach is in the following main characteristics:

- It would allow for more flexibility in sampling projects for evaluation during the evaluation planning stage. Evaluators could select smaller projects which might be seminal to the next generation of programming, while less significant in budgetary terms, where noted by the program for providing critical positive (or negative) lessons learned;
- It would maintain the use of rating scales to be applied for each criterion used in assessing each project but would replace the current numerical system with a simplified nominal system

with four categories. Such a “traffic-light” system has the virtue of relative simplicity, while still applying a structured approach to the review. It has the advantage of not suggesting a spurious level of rigour in the project rating system. The new rating system would be more action-oriented in that each rating level would be defined in terms which could influence action taken (or not taken) by the program either in the current program period or in the design of the new program.;

- Option One would require the development of detailed guidelines and definitions for each evaluation criterion and for each rating at program and at project level for each criterion.

The strengths and weaknesses of option One are shown below.

Strengths	Weaknesses
<ul style="list-style-type: none"> • Privileges accountability as the highest priority for CPEs (and thus retains both centralization and the application of common criteria and methodology aspects) • Introduces elements of flexibility and simplicity (in sampling for flexibility and in rating scales for simplicity) • Creates some space (time and resources) for examining issues of strategic approach and the non-project activities of country programs • Retains quantitative results assessment as basis for VfM assessment 	<ul style="list-style-type: none"> • May not go far enough in introducing flexibility • Still requires the same relatively large investment in CPEs applied in every country of focus, regardless of where the country program is in its life cycle or what its specific context (developmental or humanitarian action) • Not appropriate for addressing program innovations or adopting a risk-based approach • Not useful for addressing issues of scaling-up or replicating program innovations

Option Two: A Variable Approach

For many countries of focus, Option Two would be exactly the same as Option One. They would be situations where the DED felt that a relatively “heavy” CPE process and investment of resources were worthwhile. The countries selected for the “heavy” or “standard” application of the model could have the following characteristics:

- Substantial investments for the bilateral (country-to-country) component of programming over the course of the country program under review (most probably countries of focus);
- A reasonably stable set of country program objectives and goals and a broadly stable strategy being applied;
- The expectation (based on a review by DED) that the available evidence base (including results monitoring data and available project evaluations) would support the standard approach to a CPE;
- An expectation that bilateral programming will continue to be stable and significant.

However, it may also be possible to identify situations in which the application of the “heavy version” CPE is less practical or relevant. This could be where:

- The strategic approach to programming and high level program goals and objectives have changed significantly during the program period so that current programs and projects have different expected results. (One way to think of this is a major re-orientation of the theory of change.);
- The significance of the program in budgetary terms either changed significantly during the life of the program or is expected to change in the near future (either by scaling down or up);
- The evidence base would not support a full, results-oriented CPE; and
- The developmental/humanitarian situation in the country is either undergoing very rapid change or is at high risk/expectation of doing so soon.

In these cases it may be appropriate to apply a CPE “light” approach that would have the following characteristics:

- Less emphasis on the evaluation of project results, since these results would be less relevant to the new or emerging program context. Where the program had collected and reported information on project results this would be summarized, but measuring the results of projects would not be the main focus of a “light” CPE;
- Significant emphasis on the appropriateness of DFATD program strategies in both the pre-existing and emerging context;
- A greater emphasis on rationale issues (In the changing context does it still make sense to invest in sector X? If so, what significant changes are needed in the CP strategic approach?);
- More focus on partnerships and the role of DFATD-supported programs in the context of work by other bilateral agencies, by multilateral agencies and by the national government and civil society;
- More focus on coordination and alignment issues and less on direct assessment of the results of programming; and
- A more rapid and flexible CPE design process in countries where a “light” CPE approach is considered appropriate.¹⁹

¹⁹ It is believed that the “light” approach would be more rapid since it would not include the same focus on project results and, as a result, would require less time for project sampling, documenting project profiles, identifying and accessing project results and/or confirming project results in the field. It would, however, still be subject to the delays in implementation that stem from CPE planning and contracting and the review process (see Section 3.4.1).

The strengths and weaknesses of option Two are shown below.

Strengths	Weaknesses
<ul style="list-style-type: none"> • Retains positive focus on accountability for program (and project) results in appropriate countries • Allows for appropriate choice of approach, depending on country type (low- or middle-income country, fragile state) • In contexts in which programming has been changing significantly, due to changes in priorities or changes in context, the differential approach allows CPEs to be more immediately useful in program planning • Allows CPEs to be more relevant for learning agency-wide lessons in appropriate programmatic responses to changing conditions • Where appropriate can directly focus on issues of innovation and scaling up and replicating positive program developments • Allows for use of a risk-based approach in appropriate program settings 	<ul style="list-style-type: none"> • Emergence of two distinct types of CPE with different levels of emphasis on accountability • Danger over time that the “light” CPEs, if they are more rapid and are seen to be more useful to the programs themselves, would crowd out the conventional “heavy” CPEs with some subsequent reduction in the measurement of results and accountability • May complicate quality assurance process given the emergence of two distinctly different types of CPEs with different quality criteria • Provides a more implicit than explicit assessment of project level results (or at least a less quantitatively precise one) which may weaken utility for VfM analysis

Option Three: Decentralized and Integrated

Option Three would be the most radical departure from current practice. It would decentralize the CPE function so it becomes a responsibility of the bilateral programs and is fully integrated into the process of developing, approving and implementing a new country program. This would in turn require a very significant re-orientation of the role of the Development Evaluation Division (DED) in the planning, conduct and reporting of CPEs.

The main characteristics of Option Three include:

- Less emphasis on accountability as the primary purpose of CPEs; more on learning and utilization for country program design and implementation, though would still be self-reporting on accountability by the program;
- Retention of significant elements of standardization including a set of core evaluation issues to be addressed;
- Flexibility for the country program to introduce supplementary issues and to sample selected strategic projects with implications for future programming;
- If deemed necessary, use of a nominal rating for results at the project and program levels with common guidelines and definitions of ratings;

- DED in an advisory and quality assurance role through the development of core issues, core criteria and guidance on sampling, ratings definitions and guidelines;
- Integration of CPEs into the country program planning and approval process so that new Country Development Programming Frameworks must refer to a CPE completed after the half-way point in the current program.

The strengths and weaknesses of option Three are shown below.

Strengths	Weaknesses
<ul style="list-style-type: none"> • Raise the utility of CPEs for program planning decision-making generally, by integrating them into the process and timeline for country program planning • Places responsibility for accountability (the frequency, quality and utility of CPEs) directly with the country programs themselves so that country programs would be less able to distance themselves from the results of CPEs • Placing responsibility for both project and program evaluations with the country programs would increase the incentive for programs to acknowledge the link and to address it 	<ul style="list-style-type: none"> • Self-evaluation by the country programs themselves implies reduction in independence and impartiality. • If CPEs are funded from program budgets, this would involve a slight reduction in the portion of country program budgets available for direct programming. On the other hand, it would save the department O&M resources. • Would require increased HR capacity in program branches to plan and implement evaluations, and could result in variable quality.

Conclusion

The report's main messages include: Treasury Board and the Development Assistance Committee do allow flexibility in seeking the right balance between accountability and learning, in the use of different methodologies to assess prescribed evaluation criteria, and the inclusion of specific issues; CPEs would be more useful to programming if they would address the relevance and rationale for country programs related issues, and cover non-project activities; Ongoing data collection by the country programs of information on results is essential to improve the availability and quality of data for CPEs; Better integration of monitoring and evaluation functions is needed; Contribution analysis could be a useful tool given that CPEs are typically done in complex context; Guidance should be provided to consultants on how to apply the rating scale (if they are continued); Many factors contributing to delays of CPEs can only be addressed by working with other corporate systems; The measure of efficiency and economy can be strengthened; Pre-CPE corporate mechanisms (for instance data quality), and post-CPE corporate mechanisms (for instance knowledge management) are key to increasing the usefulness of CPEs (see Table 1).

The report suggests elements of solutions for the 15 issues identified. In addition, it presents three options in which some of the identified issues can be addressed, depending on the direction chosen. The options range from a focus on meeting senior management's needs for external accountability to a focus on meeting program improvement needs of program managers. They include different approaches to flexibility in the issues to be addressed in a CPE, the balance between the evaluation of projects and non-project activities, the nature of a project rating system, the use of a common approach and methodology for all CPEs, and the organizational arrangements for carrying out CPEs. These options do not preclude conducting evaluation of country programs with a focus on a sector, theme or programming approach, as other donors do.

Following consideration of these options and decision on a preferred direction, the Development Evaluation Division will undertake work to develop appropriate evaluation methods, data collection tools, and clear guidelines for the next generation of country program evaluations, to appropriately respond to current needs for accountability, learning and decision-making.

5 Annexes

Annex A: Rationale for Country Program Evaluation Review

In line with the DAC's quality standards for evaluations, DFATD's objectives for the review of the approach to CPEs are as follows:

- assess current processes and products for CPEs;
- assess main user needs for CPEs
- review existing methodologies and best practices in international development program evaluation;
- suggest approaches and designs with sufficient flexibility to meet user needs as well as accountability requirements

The issues for this review have been identified through internal discussions within DFATD and between DFATD and GGI staff. The issues can be grouped at two levels – strategic (what and why) and operational (how). The strategic issues focus on the overall intent and use of country-focused evaluations and how they fit within the overall global approach to development programming and evaluation. The operational issues focus more on how country program evaluations should or could be implemented, in relation to DFATD's programming, accountability frameworks and approach to conducting evaluations.

Strategic issues

1. Is there still a need for evaluations focused on DFATD's programming in a given country? Should their primary focus be for accountability or learning lessons for programming? Should they provide information for decision-making (including policy development) and, if so, who should be the primary and secondary audiences (from country to senior management levels)? Should they address DFATD's accountability requirements with the Treasury Board Secretariat?
2. How effective have recent CPEs been in providing useful information for DFATD?
3. What is the current theory about, and approach to, the design and use of country program evaluations, in light of recent changes in development programming and evaluation (e.g. Paris Declaration, use of theories of change and contribution analysis)? Are there other approaches being used by other bilateral agencies?
4. How do DFATD's current CPE approach and methodology compare with those of other bilateral agencies?

Operational issues

5. What should be the overall approach to country program evaluations? Should they have a strategic or operational focus? Should they be carried out at a centralized or decentralized level?
6. Should the CPE approach be appropriate for the range of countries in which DFATD is programming (e.g. middle income countries, low income countries, fragile states, states in

conflict, countries of focus, etc.)? How can the approach be flexible enough to accommodate different contexts? How could the current approach be more appropriate?

7. Is there a need for consistency in the approach to evaluations across countries and does the current approach and methodology allow for this? What mix of qualitative and/or quantitative methodologies would be appropriate for this?
8. How does the current CPE approach take into account other corporate management, monitoring and reporting tools (e.g. Country Performance Measurement Frameworks, Annual Country Reports) and/or Branch-led evaluations? Does it make use of these tools? How could it be improved?
9. How can the approach and methodology adequately ensure quality and consistency in its application in all evaluations? What is a reasonable range of resources (financial and time) for the evaluation to be affordable and timely?

The approach to this review involved interviews, discussion groups and a document/literature review. In addition to the formal methodologies, the DFATD project manager monitored the implementation of current CPEs to provide additional inputs to the review. Interviews have been conducted with DFATD staff, including senior and program managers, with representatives of other bilateral agencies, and external members of DFATD's Evaluation Committee. Discussions groups have been held with consultants who worked on country program evaluations for DFATD over the last five years.

Annex B: Bibliography

- American Evaluation Association. 2011. The Program Evaluation Standards. <http://www.eval.org/EvaluationDocuments/progeval.html>
- Bill and Melinda Gates Foundation. 2010. A Guide to Actionable Measurement. Seattle, Washington.
- _____. 2011. 'How Do We Make Measurement Core to Our Business?' 20 September. <http://www.impatientoptimists.org/Posts/2011/09/How-do-we-make-measurement-core-to-our-business>
- Canada School of Public Service. 2013. Evaluating Federal Programs and Initiatives – Introduction. Participants Manual. Government of Canada.
- Canadian International Development Agency. 2012. Draft Evaluation Policy.
- Care International. Care International Evaluation Policy. www.care-international.org/.../82-CI-Evaluation-Policy-English-condensed
- Centre for Disease Control and Prevention. 2012. Office of the Associate Director for Program Evaluation. September. <http://www.cdc.gov/eval/framework/index.htm>
- Conway, T and Simon Maxwell. 1999. Country Program Evaluation: A State of the Art Review. In OECD, Development Assistance Committee (1999), Evaluating Country Programs. Vienna Workshop 1999. Paris.
- Cosée, O. 2001. A Review of Country Program Evaluations in Selected UN Organizations and International Banks. *Commissioned by UNICEF Evaluation Office*. 15 March.
- Dahlgren, Stefan. 2007. Evaluations of Country Strategies: An Overview of Experiences and a Proposal. *Sida studies in Evaluation*. April.
- Evaluation Cooperation Group. "Good Practice Standards: Country Strategy and Program Evaluations", Operations Evaluation Department, Asian Development Bank, 2008
- Flint, Michael, John Gray, and Stephen Jones. 2004. Evaluation of DFID Country Programs – Synthesis Report. Produced for *Department for International Development*. July.
- Governance and Social Development Resource Centre, Research Report on Value for Money, September 2010.
- Independent Commission for Aid Impact, ICAI's Approach to Effectiveness and Value for Money, 2011.
- International Fund for Agricultural Development. 2009. Evaluation Manual: Methodology and Processes. April.
- James Irvine Foundation. 2011. The James Irvine Foundation Evaluation Policies and Guidelines. <http://irvine.org/evaluation/our-approach>
- Keijzer, N. and J. Oppewal (2012) "Learn to walk before you run? Discussion paper on methodological approaches to evaluating coherence in international cooperation." *ECDPM Discussion paper* No 132, June. European Centre for Development Policy Management.
- McNamara, Carter. 2006. Field Guide to Nonprofit Program Design, Marketing and Evaluation. Authenticity Consulting, LLC. <http://managementhelp.org/evaluation/program-evaluation-guide.htm>

- Ministry of Foreign Affairs of Denmark. 2006. Danida Evaluation Guidelines. Evaluation Department, *Ministry of Foreign Affairs of Denmark*. November.
- Morra Imas, Linda G. and Rist, Ray C. *The Road to Results: Designing and Conducting Effective Development Evaluations*, The World Bank, 2009
- OECD Development Assistance Committee. 1991. Principles for Evaluation of Development Assistance. Paris. www.oecd.org/dataoecd/31/12/2755284.pdf
- Sasaki, Ryo. 2012. An In-Depth of International Comparison of Major Donor Agencies: How Do They Systematically Conduct Country Program Evaluation? *Journal of MultiDisciplinary Evaluation* 8 (18).
- UNICEF. 2001. A Review of Country Program Evaluations in Selected UN organizations and International Banks.
- United Nations Development Program (UNDP). 2004. Report of the Workshop on Assessing Development Results at Country Level. Report from a conference in New York, November.
- United States Department of Housing and Urban Development. 1978. Program Evaluation and Analysis: A Technical Guide for State and Local Governments. Washington, DC. https://www.bja.gov/evaluation/guide/documents/step_4.htm
- World Bank (Office of Evaluation). 2003. Toward Country-Led Evaluation. A Multi-Partner Evaluation of the Comprehensive Development Framework.
- World Bank (Office of Evaluation). 2005. Country Assistance Evaluation Retrospective: OED Self-Evaluation.
- Zurcher, D and P. Walther. 2006. Country Program Evaluations, Lessons Learned from Danida, DFID, SDC and Sida. Mandated by SDC. Zurich, June.

Annex C: Summary of Generic CPE Terms of Reference

1. Evaluation Rationale and Purpose

In accordance with the Federal Accountability Act and the Treasury Board and Agency's Evaluation Policies, the evaluation will have learning, dissemination and accountability purposes, namely:

- To assess the Country Program's overall performance in achieving results from fiscal year XX to YY).
- To provide DFATD with relevant information for future programming.

2. Evaluation Object and Scope

The evaluation will cover a five-year programming period (specify fiscal years) during which time the Country Program was guided by the following strategic documents:

- Country Strategy (title and period covered)
- Country Development Programming Framework (title and period covered)
- List other strategies/frameworks if the period to be reviewed overlaps several programming cycles.

The evaluation will focus mainly on the disbursements of the Country Program that is financed by DFATD's Geographic Program Branch (GPB), but will also review disbursements from the Partnerships with Canadians Branch (PWCB).

The evaluation will focus on the Country Program's three main sectors of disbursements during the period under review, namely sector xx (...%), sector yy (...%), and sector zz (...%). Regarding the disbursements that were given the sector coding "other", it should be noted that for the purpose of this evaluation it may be possible for the Consultant, in consultation with DFATD staff from headquarters and the field, to align some of these disbursements under the sectors to be examined by this evaluation.

3. Evaluation Process

Consultant's Mandate

The work of the Consultant team will be overseen by, and under the responsibility of a Lead Consultant acting as principal evaluator, assisted by a team of Canadian and/or local evaluation consultants acting as sector specialists. The Consultant will execute the mandate described below:

- complete the list of key documents (including the sources of information) required to assess the overall Country Program and each of the projects in the sample;
- review all the key documents gathered for the Country Program and for each project in the sample;
- prepare a draft evaluation Work Plan including the following:
 - a. Any changes (from the TORs) to the objectives, methodology, and evaluation issues

- b. Revised project sample to minimize sample bias, including analysis of evaluation coverage, and justification for inclusion of any projects from the less prominent sectors
 - c. Evaluation indicators tailored to address the characteristics of this specific Country Program and project sample
 - d. Possible synergy with the evaluations of other partners, if applicable
 - e. Validation of key evaluation issues (based on those identified during the scoping mission)
 - f. List of key documents gathered
 - g. Draft list of interviewees
 - h. Draft interview guides and questionnaires tailored to the needs of various stakeholder groups
 - i. Details on data collection and analysis, including the site visits (calendar, itinerary, level of effort, and direct expenses for each Consultant team member)
 - j. Details on the case studies, if applicable (including mission agenda, background materials on projects and partners)
 - k. Revised calendar of activities
- review and finalize the evaluation Work Plan based on feedback from DFATD;
 - conduct data collection in Canada and in the field;
 - prepare the Project-Level Assessments and the Program-Level Assessment, in which the Consultant will (i) describe in narrative form the results and performance of the sampled projects and the Program based on this analysis, and (ii) rate the sampled projects and the Program against each evaluation criteria (see section 2.2.5 on Criteria), using the rating scale (see section 2.2.6 on Rating);
 - provide DFATD's Evaluation Division with access to the evidence gathered during the evaluation in support of each finding, in a format agreed to by Evaluation Division – this could entail an evidence binder, or use of software (evidence will be kept in a protected file);
 - debrief DFATD (in the field and at headquarters) on the preliminary evaluation findings;
 - draft the Technical Report (using the Table of Contents provided);
 - participate in a meeting with DFATD representatives to discuss and validate the draft report;
 - review and finalize the Technical Report based on feedback from DFATD's Evaluation Division and Evaluation Committee. This should include the Consultant's response to DFATD's comments on the draft Technical Report.

3.2 Evaluation Criteria / Questions

DFATD has adopted the standard criteria recommended by the Development Assistance Committee (DAC) of the OECD for program evaluations. This evaluation will also assess two key questions (one on development results, the other on management factors) that DFATD's

Evaluation Division has set as a standard for its program evaluations. The evaluation questions and criteria are defined in table 1 (below).

Table 1: Key Evaluation Questions and Definitions for the Evaluation Criteria

What has been achieved? (Development Results)
<ul style="list-style-type: none"> • Relevance¹: The extent to which the objectives of the development intervention are consistent with beneficiaries’ requirements, country needs, global priorities, and partners’ and donors’ policies, the appropriateness of development interventions in a given sector, region or country. • Effectiveness¹: The extent to which the development intervention’s objectives were achieved, or are expected to be achieved, taking into account their relative importance. • Sustainability¹: The extent to which there is a probability of continued benefits from a development intervention after major development assistance has been completed. The resilience to risk of the net benefit flows over time. • Cross-cutting themes: The integration of the crosscutting themes across all projects and the overall Country Program: <p>5.1.1.1 Gender equality: The extent to which the development intervention has achieved results in gender equality and promoted women’s empowerment</p> <p>5.1.1.2 Environmental sustainability: The extent to which the development intervention has achieved results in environmental protection and sustainability</p> <p>5.1.1.3 Governance: The extent to which the development intervention has achieved results in democratization, human rights, the rule of law, public-sector capacity building, conflict prevention</p>

How and why were the intended results achieved or not? (Management factors)
<ul style="list-style-type: none"> • Coherence: (i) External coherence will look at coordination activities and policy dialogue with national and other international actors. (ii) Internal coherence will examine the role of the DFATD Program in coordinating DFATD’s delivery channels and Canada’s whole of government efforts. • Efficiency¹: A measure of how economically resources/inputs (e.g. funds, expertise, time) are converted to results. Given that cost-efficiency and cost-effectiveness may be complex and difficult to assess at the program level, the CPE will review the delivery systems (e.g. decentralization, Program Support Unit) and mix of resources (e.g. local, technical resources) at headquarters and in the field. • Management Principles²: A measure of the application of the Paris Declaration principles of ownership (i.e. developing countries set their own strategies for poverty reduction, improve their institutions, tackle corruption), alignment (i.e. donor countries align behind these objectives and use local systems), and harmonization (i.e. donor coordinate, simplify procedures, share info to avoid duplication). • Performance Management: A measure of the development intervention’s management strategy for assessing the performance of development interventions against stated results (outputs, outcomes, impacts), including results-based management (RBM) and monitoring/evaluation functions, risk management, and timely corrective action to address problems when they arise. • Risk Management: To what extent (i) was there a system in place prior to project or Program implementation to identify, monitor, report, and manage operational, development, reputational and financial risks; (ii) was the system used, (iii) have the outlined risks affected the project or Program; (iv) were the risks effectively mitigated.
1 OECD-DAC, Glossary of Key Terms in Evaluation and Results Based Management, OECD, 2002
2 OECD-DAC, Paris Declaration on Aid Effectiveness, OECD, Paris, February-March 2005

3.3 Sampling

The evaluation universe includes ## different projects and \$\$\$ million in Canadian Official Development Assistance (ODA). Reviewing 100% of these projects would be difficult and costly. Given the degree of variability of DFATD projects and partners, a randomly selected sample might leave out key investments, disproportionately misrepresent different-sized partners, and affect the usefulness of the evaluation. Therefore, a stratified sampling methodology based on the characteristics and specificity of the Country Program will be applied (Table 2 below).

Table 2: Sampling Criteria

Sampling Criteria	Description
Sector / thematic coverage	The projects selected (usually about 25) represent at least half (50%) of the Program’s total budget and of the sector or thematic value of the Program.
Delivery modalities	The evaluation will not analyse “core funding” from PWCB or MGPB (these will be covered in other DFATD evaluations). The

Sampling Criteria	Description
	<p>sample should include, as appropriate:</p> <ul style="list-style-type: none"> Multi-donor investments Program-Based Approaches (PBAs), including general or sector budget support, pooled funds, trust funds, etc.; Bilateral “directive” and “responsive” projects, including “multi-bi” projects where bilateral funds are channelled through a multilateral organization; PWCB responsive interventions.
Channels of cooperation	In each sector, a selection of bilateral GPB investments should be included. If possible, a selection of PWCB investments should be added.
Evaluability	The sample will only include investments for which sources of information are available and accessible in a cost-effective way.
Materiality	The evaluation will only include investments that are substantive enough and have a minimum approved budget of \$250,000.

3.4 Data Collection

DFATD’s “Country Program” evaluations have limited resources and as such cannot undertake detailed “project-level” evaluations, but instead must rely largely on secondary sources of information. Therefore, the evaluation will depend highly on the quality and quantity of evidence-based information that is accessible from different reliable sources.

The Consultant may use various data gathering methods, such as (but not exclusively) those outlined in table 3 (below), to obtain multiple lines of evidence. The Consultant is invited to propose its own approach to this evaluation.

Table 3: Data-Gathering Methods

Methods	Description
Document review	<ul style="list-style-type: none"> DFATD Program-level documents: Country Strategy, Country Development Programming Framework (CDPF), Logic Model (LM), Performance Management Framework (PMF), Gender Equality Strategy, Strategic Environmental Assessment (SEA), Risk Register, Annual Project/Program Performance Reports (APPR), Annual Country Reports (ACR), Program Monitoring and Reporting Tools (PMRT), Country Overviews (from PWCB), previous studies/evaluations/institutional assessments, etc. DFATD project-level documents: Project approval documents, Logic Model (LM), Performance Management Framework (PMF), monitoring reports, end-of-project reports, Annual Project/Program Performance Reports (APPR), Gender Equality Strategies, Strategic Environmental Assessments (SEA), Management Summary Reports (MSR), Investment Monitoring and

Methods	Description
	<p>Reporting Tools (IMRT), trip reports, previous studies/evaluations/institutional assessments, etc.</p> <ul style="list-style-type: none"> • <u>Documents from credible sources</u>: Other evaluation products (see the Development Evaluation Resources Centre (DEREC) on the DAC site), surveys, organizational assessments, and focus group reports. Country or sector reports, studies, and statistics. Whenever possible, gender-sensitive or sex-disaggregated data will be used.
Meetings with key people	<ul style="list-style-type: none"> • <u>DFATD managers and staff</u>: responsible for the Program or for projects, including DFATD specialists and locally engaged professionals, at headquarters and in the field. • <u>Other Canadian Government Departments</u>: as appropriate. • <u>Partner country officials</u>: at the national or local levels involved in the DFATD Program. This may include representatives from key local beneficiary organizations, depending on the level of effort and within the limits of available resources. • <u>Representatives from partner organizations</u>: Canadian Executing Agencies and regional, international or locally based partner organizations, including local consultants, involved in the management and implementation of specific initiatives. • <u>Representatives from the international community</u>: international financial institutions, United Nations agencies, multilateral organizations, DAC members, etc. <p>The Work Plan will describe the approach to meet key people and the proposed tools (interview guides, questionnaires, etc.) tailored to the needs of the stakeholder groups.</p>
Site visits	<ul style="list-style-type: none"> • <u>Visits to projects and institutions</u>: a limited number of site visits will take place in the field to complement and validate the secondary sources of information. <p>The Consultant will consult with the DFATD Program, the Head of Cooperation, and the Evaluation Division, to determine which key projects/institutions may be visited.</p>
Case studies	<p>Certain projects or methods of implementation/programming could be considered for case studies, depending on the level of effort and within the limits of available resources.</p>
Roundtables	<p>Roundtables may be held to get feedback and keep an open dialogue with the DFATD Program managers and partners at key points during the evaluation process, such as:</p> <ul style="list-style-type: none"> • At the beginning of the exercise to introduce the Consultant and launch the process.

Methods	Description
	<ul style="list-style-type: none"> • During the site visits, for key stakeholders (e.g. international community), if needed. • At the end of the site visits, to present preliminary issues/findings that have emerged to the DFATD teams in the field and at headquarters, and to discuss next steps. • At other key points, for example to validate the Technical Report's findings/conclusions, or discuss the Synthesis Report's recommendations, etc.

3.5 Rating

To ensure comparability, the Consultant will systematically apply the same evaluation questions, criteria, rating scale, and assessment grid to all of the projects selected in the sample. The assessment grid must include the requisite elements to assess performance for each crosscutting theme separately (gender equality, environmental sustainability, and governance).

The Consultant will rate the performance of each sampled project (in relation to each evaluation criteria) using both the numerical and nominal five-point scale shown in table 4 (below). The scoring must be supported by the qualitative and/or quantitative evidence gathered through the various data-gathering methods, and be consistent with the analysis and conclusions presented in the Project-Level Assessments.

The Consultant will follow the same approach for the Program-Level Assessment, but in rating the performance of the Country Program (in relation to each evaluation criteria) the Consultant will use only the nominal five-point scale shown in table 4 (below).

It is important to note, that while projects will be "assessed", recommendations will not be made for specific projects. However, the project-level assessments and ratings will inform the Consultant's sector analysis in the evaluation report.

Table 4: Program and Project Rating Scales

Nominal Scale (Program & projects)	Definition	Numerical Scale (projects only) choice of 3 values
Highly Satisfactory	The project/Program meets all the evaluation indicators for the given criteria	4.1 – 4.5 – 5.0
Satisfactory	The project/Program meets the main evaluation indicators for the given criteria	3.1 – 3.5 – 4.0
Partially Satisfactory	The project/Program meets some of the evaluation indicators for the given criteria	2.1 – 2.5 – 3.0
Unsatisfactory	The project/Program does not meet the main	1.1 – 1.5 – 2.0

Nominal Scale (Program & projects)	Definition	Numerical Scale (projects only) choice of 3 values
	evaluation indicators for the given criteria	
Highly Unsatisfactory	The project/Program does not meet any of the evaluation indicators for the given criteria	0 – 0.5 – 1.0
n/a	Cannot be assessed or not available	n/a

Table 5: Compilation Table – Summary of Project Ratings by Criteria and by Sector

#	List of Sampled Projects (for Sector XX)	Relevance	Effectiveness	Sustainability	Gender Equal. (crosscutting)	Env. Sustain. (crosscutting)	Governance (crosscutting)	Coherence	Efficiency	Management Principles	Performance Management
1	A012345-001 Project Name (and acronym)	4.0 Satis.	4.5 Highly Satis.	2.5 Partial Satis.	1.5 Unsat.	1.0 Highly Unsat.	n/a	4.0 Satis.	4.5 Highly Satis.	2.5 Partial Satis.	2.5 Partial Satis
2	etc... etc...										
3	etc... etc...										
	Average Score (for Sector XX)										

The Consultant Team is expected to build-up internal consensus about the application of the scale and ensure that all team members share a common understanding of utilized concepts (what does “highly satisfactory” really mean?). It should be the Consultant team member, not individual consultants, which make final decisions about ratings. A peer review process (among evaluation team members) for every rating contributes to prevent potential biases.

Annex C : Approaches of Other Donor Agencies

Country program evaluations first emerged following seminal aid effectiveness meetings (i.e. Paris Declaration), which encouraged donors to shift towards country-led development and program-based approaches. Country Strategies became a dominant programming tool. Donors (both bilateral and multilateral) then began to develop methods to evaluate their country programs. Most donors began to adopt CPEs in the 1990s.

Since then, a majority of donors used the OECD/DAC's criteria for their evaluations, although they tend to interpret criteria differently. Many donors have included additional criteria to better suit their needs (Zurcher-Walther, 2006). Donors also used several approaches to their evaluations, although the scope varied significantly. Some donors, such as Australia, tended to concentrate on the role of country strategies in framing assistance, but had limited focus on the institutional dimensions of country program management and produced only indicative findings about effectiveness. They were weighted towards recent, current and near-future activities. This approach contrasts with the more ambitious country program evaluations conducted by the World Bank and the Asian Development Bank, which look at impact, examine both the majority of country program components and overall country program performance, and address a long period of donor-partner relations (Sasaki, 2012).

Two key reports provide the most comprehensive overview of donors' experiences with country program evaluations. The first was conducted in 1999 by the Overseas Development Institute, in preparation for a high-level OECD-DAC conference in Vienna (Maxwell and Conway, 1999). The second, undertaken five years later, compared and contrasted the UNDP's experience with CPEs relative to that of other development agencies (UNDP, 2004). These two reports, with other (less extensive) analyses of donor CPEs during that period highlighted the following persistent challenges (Dahlgren, 2007; Flint et al, 2004; Cossée, 2001; Zurcher-Walther, 2006, World Bank, 2005; Sasaki, 2012):

- Attribution: difficulty of assigning the contribution of a specific aid agency to development results in a country;
- Usefulness: inward-looking and aimed at a limited range of stakeholders; trying to cover all the bases with limited time and space for detailed analysis; conclusions and recommendations too general;
- Design: overly long and broad evaluation designs, with too many projects included in the sample;
- Timing: limited connection between evaluations and the program planning process (relative to other monitoring exercises);
- Consistency: inconsistent interpretations and conclusions based on subjective assessments (rather than numerically measurable data);
- Overlap: lack of distinction between end-of-cycle reviews and country program evaluations;
- Ownership: limited sense of ownership and engagement by country offices and partner countries;

- Data quality: limited availability and quality of data to evaluate since some country programs are not developed with a plans for how goals and objectives will be measured at the end of the cycle – a problem that stems from the lack of a results focus in programming;
- Cost: high cost of undertaking regular CPEs;
- Clarity: if not clearly and diligently explained, the complex criteria and methods used in CPEs can lead to misunderstanding of findings and conclusions; and
- Ratings: inconsistencies between ratings and narratives.

According to Dalhgren (2007), many multilateral organizations use country program evaluations as a basis for planning future country programs, in spite of the challenges listed above; although what is included in CPEs varies considerably between organisations. On the other hand, there are many bilateral donor countries, such as the Netherlands, Norway, the UK, Sweden and the United States, which have stopped conducting CPEs because they are too expensive and complicated in relation to their usefulness.

Multilateral regional development banks have indicated that, given their country-based membership and their governance system, it makes sense for them to focus on country strategy/program evaluations. Bilateral aid agencies from the Netherlands, UK, Sweden, USA, Australia and France also indicated that they have stopped doing country program evaluations. Many European aid agencies prefer to work closely with the European Commission. Requirements of the U.S. Congress on USAID make CPEs less relevant. Differences in governance systems and accountability requirements influence how bilateral aid agencies structure their evaluation agenda.

The Netherlands has only conducted one country program evaluation in the last 14 years. Denmark covers specific sectors or themes for a specific country (as opposed to full country program evaluations), often covered through joint evaluation with other donors and/or partner country. Norway only does country program evaluations on an as needed basis. They tailor the terms of reference for each, keeping in mind the country context and the structure of the Norwegian support to the country.

The United Kingdom's Department of Foreign Affairs and International Development (DFID) no longer carries out country program evaluations. This role was taken from DFID with the establishment of the Independent Commission for Aid Impact (ICAI) in 2011. ICAI's studies are mostly reviews and audits rather than evaluations *per se*. ICAI could do a country program evaluation study, but this does not feature in their current work plan. DFID's country offices are now expected to take forward their own evaluation work, including overall country program evaluations, and some work has been initiated on how this could be done in a consistent way across the organisation. DFID's role now is to evaluate individual projects, policies and thematic programs. However, these are commissioned and led by the operational units in DFID themselves, rather than by the central evaluation function.

USAID has done some CPEs in the past. The Agency now prefers instead to look at specific program components across many country programs, or look at one given sector across many country programs. Very rarely, will evaluations encompass a complete look across a whole program and this would more likely be done through an audit than an evaluation. USAID does not consider CPEs useful. For accountability purposes they do "management assessments," which focus on resource management. For learning purposes they focus on program components or on sectors/themes.

USAID and DFID's respective evaluation services provide support to country offices as these now lead their own country program evaluations, based on their perceived needs. In doing so, the two agencies consider they can focus better on corporate evaluations and that they can increase the frequency and consistency of interacting with stakeholders to enhance buy-in and ownership of evaluation results.

AusAID has only completed three country program evaluations since 2006, with a further two ongoing. This agency however, is currently considering the use of country program evaluations in a more systematic mode. It seeks a more comprehensive consideration of outcomes over a period of assistance that is long enough to provide concrete evidence of development results, in addition to a more robust consideration of the institutional factors that have influenced the effectiveness of country program assistance.

France's AFD (Agence française de développement) does few CPEs and only jointly with the French Foreign Affairs and Finance departments. Occasionally, AFD would do joint CPEs with the European Commission. AFD has mixed feelings about these joint evaluations, given the complexity of coordination. AFD is currently considering doing reviews (not full evaluations) of its country strategies, including greater involvement of recipient countries in such reviews.

Another observation from Dahlgren (2007) is that no organisation among multilaterals and bilateral agencies appears to have the ambition to evaluate country programs as a whole "probably because nothing like that exists." (Dahlgren, 2007, p. 41). Some organisations have an interest in examining what could be called the 'internal effectiveness' of the organisation's country program, but few organisations appear to be interested in cost-effectiveness at country program level, given the complexity of the process and its limited value added to decision-making. Also, most bilateral donors have moved toward identifying the contributions their agencies have made to development outcomes. In fact, some (e.g. USAID and DFID) suggest that the most accurate description of their contribution to development is building enabling environments and conditions for a country to pursue development interventions.

SIDA, USAID, AFD indicated that they prefer a 'light approach' in the formulation of their evaluations in general. As a general principle, these agencies feel that evaluations should be narrowly focussed, based on existing material, take little time and involve few consultants.

Bracegirdle (2002) confirms that bilateral donors have been using both light and heavier program evaluations for some time, in response to a range of conditions and factors. Some program evaluations are large, expensive initiatives lasting two to three years. Others are much smaller in scope and program coverage, and designed to make an immediate contribution to the donor's programming cycle. Generally, the exercise is 'lighter' or 'heavier' depending on the structure of the country program, the evaluation issues being addressed, the amount of programming covered, the time-period under review, and the forms of the donor-partner relationship being examined. While 'heavier' program evaluations produce more conclusive findings, and allow more comparative review of country program performance, 'lighter' evaluations can produce policy-relevant recommendations rapidly, which is particularly useful when the aid program is set within changing circumstances.

Light version of program evaluations typically cover selected program components, a limited selection of evaluation criteria (often: relevance, efficiency, coherence, effectiveness), and cover a short period of programming. Also, a light evaluation can be designed to focus on program-level matters only (policy dialogue and aid effectiveness practices, for instance) and not include the assessment of projects.

Like bilateral donors and International Financial Institutions, United Nations agencies have shown differing levels of commitment to the concept of country program evaluation and this commitment has varied over time within the agencies concerned.

At UNICEF, for example, country offices and program departments are still able to commission and carry out CPEs. They may request support in this activity from the Evaluation Office at headquarters (up to and including having the Office carry out the evaluation). In practical terms however, the practice of country program evaluations has fallen from favour at UNICEF. A review of UNICEF evaluations carried out in 2012 found that only one country program evaluation had been carried out in the previous three years.²⁰

In contrast, at UNFPA, the practice of CPEs has recently been revitalized with a new evaluation policy, a re-arrangement of the lines of authority for country program evaluations and a new set of guidelines.²¹ CPEs at UNFPA are a compulsory part of the program planning cycle and must be completed before a new program can be approved.

UNDP continues to invest in regular CPEs although a review of their evaluation reports conducted in 2012 noted a decline in frequency over the prior four year period.²²

The Inter-American Development Bank has an interesting typology of country assistance evaluations, each type addressing specific needs: country studies (broad assessment of the political, economical and social situation of a country); country strategy evaluation (an assessment of the relevance and effectiveness of the Bank's strategy in a given country); country program evaluation (an assessment of the relevance, effectiveness and efficiency of country-level operations); country assistance evaluation (an assessment that includes program sustainability and contribution).

The key concern for aid agencies remains the amount of resources and time they should spend on CPEs. Also, needs for information differ greatly from one country program to another. For most multilateral and bilateral donors, a standard approach applied to all program evaluations is not good practice anymore. Corporate evaluation services have diversified the design of their evaluations to address a narrower set of issues, within a wider variety of programs, both in size and in nature.

²⁰ "Review of UNICEF's Development Effectiveness 2009 - 2011: Final Report", OECD-DAC, 2013

²¹ "Handbook: How to Design and Conduct a Country Program Evaluation at UNFPA", UNFPA, March, 2012.

²² "Development Effectiveness Review of the United Nations Development Program (UNDP) 2005-2011", Canadian International Development Agency

Annex D: Policy Dialogue

Policy dialogue is essential for aid effectiveness. Local ownership requires policy dialogue, often because there is a need to support (or encourage) shifts in the policy priorities and programming approaches of our partners. As policy dialogue implicates all stakeholders within the aid architecture, it is directly related to harmonization (in terms of shared instruments like PBAs) and mutual accountability for development outcomes. Policy dialogue is also important for effective programming. There is a strong feedback loop between policy dialogue and outcome achievement for aid initiatives.

Most aid agencies do not measure the impact of policy dialogue and do not document their policy dialogue activities, nor do they have specific guidelines or models. Many, as the World Bank and DFID, encourage their staff to monitor their success in advocacy as part of the self-assessment process, but formally, there is no procedure for tracking the outcomes of policy dialogue.²³ DFATD has developed a set of e-lessons learned based on a meta-synthesis of a number of DFATD evaluations and reports.²⁴ Some programs have done interesting work about the use of policy dialogue to improve programming.²⁵

Donors conceptualise the process and the result of dialogue differently. AusAID defines the process as ‘working with partner countries’ and the result as ‘accelerating growth’. The United States Agency for International Development (USAID) understood the process of dialogue as:

... a mechanism to incorporate the interchange of ideas and information whereby either viewpoints or both can change to bridge the initial differences between the two.
[Source?]

And the result as:

the aid recipient comes to view the policy advice as genuinely in the interest of its own economic progress

Dialogue between donors and recipient countries often appears to be more about the allocation of resources (for example, dialogue over budget support), but is less often guided by values. The United Nations Research Institute for Social development proposes a definition that brings the politics to the forefront:

Policy dialogue is defined as organized deliberation between two or more actors on the allocation of values that is likely to result in new policies or modification of existing ones.

Here, policy dialogue has the possibility of transformational outcomes rather than transactional outcomes. The negotiation of the allocation of values should not be understood as one-off events,

²³ AusAID, Review of Literature and International Practice in Policy Dialogue, Policy Dialogue Evaluation, September 2011.

²⁴ See the [Policy Dialogue/PRB Learning Series](#): “What We’re Learning About Policy Dialogue,” DFATD, March 2002. And the e-lessons database, [Information Sources: Policy Dialogue](#) on Entrenous (Evaluation/Key Resources/others/e-lessons/subject search/policy dialogue).

²⁵ The Vietnam, Ghana, Mozambique and Rwanda programs have all flagged the importance of policy dialogue in their programming. See: [EDRMS# 5657328 - DRAFT VIETNAM POLICY INFLUENCING STRATEGY](#); [EDRMS # 5906024 Reporting on Mozambique Policy Dialogue Results](#); [EDRMS #5980570 Ghana Policy Dialogue Objectives: How to Measure Progress?](#); [EDRMS # 5980565 - Rwanda Policy Dialogue Engagement Plan](#)

but rather as a continuous process which occurs at the initial political stages of policy change, through more institutional aspects of policy implementation to operational policy outcomes.

Measuring the impact of policy influencing work presents a particular set of challenges. First, there is the difficulty in determining the links between policy influencing activities and outputs and any change in policy. While some policy changes can be fast and have easily detectable impacts, most policy change is incremental, highly complex and often proceeds in a nonlinear and unpredictable fashion with its process shaped by a multitude of interacting forces and actors. This makes it almost impossible to predict with confidence the likely consequences of a set of activities on policy. Second, success is somewhat difficult to define and may be altered along the way. Policy changes tend to occur over long timeframes that may not be suitable to measurement in the usual rhythms of projects and evaluations in aid agencies.

Best practices for assessing policy dialogue outcomes for partner institutions and organizations include the following:

- **Realistic Theory of Change (ToC):** based on a clear understanding of the context and drivers²⁶ for change within the target organization or individual(s), develop a realistic ToC for the initiative or program;
- **Policy Dialogue Strategy:** develop a strategy that includes clarity on *what* the program wants to achieve within a particular policy space and *how* to achieve results through the dialogue process (i.e. prioritization of our key influence activities within the aid architecture), and inclusion of this in the program or sector-specific work plans would permit ongoing monitoring and reporting;
- **Embed policy outcomes (and indicators) in the Logic Model (LM):** ensure that the LM includes outcomes with policy content focusing on actors, processes and institutions, along with fairly sensitive indicators to measure change for regular monitoring and, where required, adjustment to the Performance Management Framework;
- **Measuring Influence:** as policy change is based on actors, the relationships between them and the institutions in which they work, it is important to keep systematic track of the various actors, their interests, ideologies, capacities, their alignment with program goals and their relationships with other players - and - how all these change;²⁷
- **Regular assessment:** policy change, being a highly complex process involving many stakeholders both domestic and international, requires a ToC that is flexible and iterative. Best practice suggests a regular assessment by the programming team, including revisiting of the ToC, questioning any underlying assumptions and appraising relevance of expected outcomes; and

²⁶ Stachowiak, S, "[Pathways to Change: 6 theories about how policy change happens](#)", Organizational Research Services, 2009.

²⁷ Jones, H, [A guide to monitoring and evaluating policy influence](#), ODI, February 2011, p. 9. See also: Reisman, J, Geinapp A, and Stakowiak, S, [A Guide to Measuring Policy and Advocacy](#), Organizational Research Services, 2007. See also: Start, D and Hovland, I, [Tools for Policy Impact](#), RAPID, ODI, October 2004.

- **Reporting:** Provided the ToC for policy outcomes is explicit in the LM, then reporting against these outcomes, using data from the indicators, is a fairly straight-forward exercise. Reporting on policy outcomes also needs to address the political and policy context, the nature of the evidence, key actors and the relationships and networks between them, as well as external factors. Development of a “log book,” would be a good way to track policy dialogue activities.

Australia has developed a series of hypothesis to assess the success of policy dialogue.²⁸ Policy dialogue between AusAID and country counterparts is likely to be successful if all of the following factors are in place:

- i. AusAID and its counterparts approach the dialogue with reasonable clarity of intent and reasonable expectations.
- ii. AusAID and its counterparts have comparable negotiating capital—values are balanced, and all parties engage on relatively equal terms and with equally good information and analysis.
- iii. AusAID staff and their counterparts have a shared and credible evidence base on which to base their engagement.
- iv. The fora in which dialogue takes place are neutral and provide space for formal and informal, technical and political engagement.

The success of policy dialogue between AusAID and its counterparts should be measured in terms of both *process* and *results*:

1. **Process:** successful policy dialogue entails a sustained interchange of ideas, perspectives and analysis between AusAID and relevant counterpart stakeholders, conducted in such a way that the process:
 - promotes mutual trust and confidence between parties
 - is focused on a clearly-defined purpose / end-game
 - generates an understanding of each party’s genuinely-expressed values
 - incorporates evidence
 - recognises the political as well as technical dimensions of policy reform, such that the conclusion represents a satisfactory outcome for both parties.
2. **Results:** successful policy dialogue must deliver results. It must put key policy issues on the agenda, yield tangible changes in policies, improve their implementation and promote sustainable development. Successful policy dialogue therefore tends to range across three stages: agenda setting, policy options, and implementation.

²⁸ AusAID, Theory of Policy Dialogue Success, Policy Dialogue Evaluation Working Paper, September 2011.

- The agenda setting stage focuses on forging an agreement on what issues require attention and outlining what broad changes are desirable.
- During the policy options stage, AusAID and its counterparts explore existing and prospective policy options and decide on which of these is feasible and appropriate, on the basis of evidence, in terms of the realities of the political economy, and in terms of the resources available.
- During the implementation phase, dialogue focuses on facilitating and tracking progress on institutional and regulatory reform and follow-through. It needs to be outcome-based – that is to say, at the policy level, this phase is about the extent and quality of change to people’s lives and livelihoods resulting from changes to the institutions and processes of government under the new policy.

An additional dimension—forming and maintaining relationships—is central to each of these stages of policy dialogue. In the first instance, this may be more about building trust and legitimacy, and gaining an understanding of the others’ priorities and intent, than about changing agendas or policies.

Success can be achieved at each stage of the dialogue process. Working with counterparts to get and keep a key development issue on the government’s agenda, for example, is an indicator of successful dialogue at the agenda-setting stage. However, longer-term success of sustained policy dialogue usually entails achieving results in all three stages.

Annex E: Value for Money

The OECD defines value for money (VfM) as the optimum combination of cost and quality to meet the user's requirements. It is usually assessed using the criteria of economy, efficiency and effectiveness. However, VfM is now being more frequently referred to as the best balance between economy, efficiency, effectiveness and equity.

The UK's Independent Commission for Aid Impact (ICAI) explains what these terms might mean if, for example, ICAI was to assess a program to deliver anti-malaria bed nets to a country with high incidence of the disease.

ICAI's Illustration of the Four Es in an Aid Context²⁹

Definition of the four Es	Application to provision of anti-malaria bed nets
Economy: getting the best value inputs	Were bed nets of the required standard bought at the lowest possible cost?
Efficiency: maximizing the outputs for a given level of inputs	Given the number of nets bought, how many people used the nets for their intended purpose?
Effectiveness: ensuring that the outputs deliver the desired outcome	For those people provided with nets, has the incidence of malaria decreased?
Equity: ensuring that the benefits are distributed fairly	Have the nets reached the poorest people and minority groups in more remote areas, as well as those closer to cities?

The four "Es" would be considered together, not separately, balancing them to come to a judgment. For example, exporting malaria nets to a country which already has local manufacturers might improve healthcare but seriously affect the economic and social well-being of people who rely on local producers for employment. Where appropriate, ICAI would consider whether the best impact is being achieved by the approach adopted. VfM is in a sense a subset of aid effectiveness. Aid effectiveness is about deciding what a program wants to do. VfM is about how to do it best. Beyond tools, VfM is a way of thinking about using resources well.

Many dismiss VfM as not relevant and impractical while others request numerical evidence of the best possible value for money in all public expenditures. Value for money is not the same as cost-cutting or efficiency. It is not about monetising everything, and cannot be assessed through only one of its dimensions in isolation. If the effectiveness of an activity is notably reduced because of cost saving, value for money is reduced. Similarly, while an activity may be very cheap and run efficiently, if it does not achieve results, it is not good value for money.³⁰

²⁹ Independent Commission for Aid Impact, ICAI's Approach to Effectiveness and Value for Money, 2011.

³⁰ Jackson, Penny, Value for money and international development : Deconstructing myths to promote a more constructive discussion, OECD Development Co-operation Directorate, May 2012.

The OECD warns that there is a danger that applying value for money could lead to a risk-averse culture in development cooperation. For instance, insisting on exact measurements (efficiency, unit costs, cost-benefit) in all projects can exclude types of projects where these things are harder to measure, encouraging a focus on things that are easier to measure rather than on what is most needed or most effective. It can also discourage innovation, since it tends toward the tried and tested types of project, with comparators and data, picked for ease of measurement rather than expected effects. Ultimately, this type of risk aversion can be very damaging to real VfM.³¹

Aid agencies across the world have adopted a range of approaches to effectiveness and value for money. Many of these approaches draw on two particular techniques for integrating these concepts into the planning, delivery and review of organizations' activities:

- **Economic analysis:** the aim of this analysis is to target resources effectively. This can be done via a cost-benefit analysis, where the economic benefits of a program are compared against the economic costs. Where benefits cannot easily be quantified, an alternative is cost-effectiveness analysis, where the costs of various approaches to achieving a given objective are compared. Economic analysis is carried out in various forms by DFID, the World Bank, various United Nations organizations, the European Commission and USAID.
- **Results-based management:** Donor agencies have made strong commitments to manage their programs in order to maximize the results they deliver. This approach aims to bring together objectives, an understanding of what needs to be done to achieve them, performance monitoring and evaluation of what has been achieved. This process is used widely, in a variety of forms, by agencies such as DFID, DFATD, the World Bank, USAID, various United Nations organizations and the Asian Development Bank (AsDB).

Many donors have adopted a range of tools to conduct results-based management. These include:

- Performance measurement framework/performance reviews (e.g. DFATD; Danida, USAID)
- Logic model (e.g. DFATD)
- Corporate results framework (AsDB)
- Results monitoring system (World Bank)
- Plan, Do, Check, Act cycle (JICA)

The adoption of a results-based management system is considered by many donors as a way to address value for money.

31 Jackson, Penny, Value for money and international development : Deconstructing myths to promote a more constructive discussion, OECD Development Co-operation Directorate, May 2012.

DFID

DFID appears to have gone the furthest among aid agencies in developing the concept of value for money. It is the only agency that explicitly uses the terminology frequently in its policies and procedures and has a Value for Money department. DFID's approach to VfM involves assessing whether level of results achieved represent good value for money against the costs incurred, moving from results to returns. Processes include the use of logframes, economic appraisals and portfolio reviews. Newer initiatives include the adoption of a business case model for project approval and the development of unit cost metrics in key sectors.

Experience from DFID shows that adding a value for money (VfM) dimension to program planning, monitoring and evaluation has more often meant making a series of informed, but subjective rather than mathematical, calculations. Applying value for money has had implications for the whole organization. Country programs have had to define clear program level objectives, in addition to measurable parameters such as acceptable timeframes and levels of risk. Successful integration of VfM depended upon good quality corporate processes and tools, and staff's acceptance and understanding of these. Value for money became expensive in financial and human resources, and complex to manage as it involved the development of complex assessment and rating systems, pressure on staff to increase compliance, better guidance, and benchmarking to ensure consistency. Corporate challenges included the availability and quality of data and the need to capture the full range of country practice. There are still concerns today over how robust the VfM analysis system is to be able to cope with a changing portfolio and the current movement towards non-traditional delivery mechanisms such as sector-wide approach (SWAs) and direct budget support.³²

USAID

USAID sees results-based management of its bilateral system as synonymous with aid effectiveness and the best way to address Congressional insistence on value for money. USAID has maintained a performance monitoring system that included: a multi-year strategic plan; an annual performance plan; and regular annual performance reports. More recently, USAID developed the "Mission Management Assessment" to track performance. Under this approach, assessment teams composed of senior officers in key operational areas (senior management, administration, finance, programming, contracting) use a peer review approach to evaluate the effectiveness of field operations and to identify best practices in both program and internal management. This approach to performance monitoring has led to a strong focus on establishing clear performance frameworks and generating data for these and a movement away from the use of evaluation as a management tool within USAID.

Sida

At Swedish Sida, some activities to entrench results-based management include:

- The establishment of a Quality Assurance Committee to review each new proposal over 100m SEK (USD 15 million) for results orientation
- Procedures for external reporting

³² Poate, Derek and Barnett, Christopher, Measuring Value for Money? An Independent Review of DFID's Value for Money (VfM) Indicator, Public Service Agreement 2003–2006, Evaluation Report EV645, Department for International Development, November 2003.

- Annual reviews of progress, rather than just a completion report (required by Sida's new three year operational planning tool). These reviews seek to allow teams to identify where and why something may be going off track or where indicators are inappropriate, and to develop remedies
- A country reporting matrix which compels country teams to identify indicators to assess progress at both country and at sector levels and to monitor them annually

Japan's JICA

While the the Ministry of Foreign Affairs (MoFA) and JICA conduct their own internal evaluations, Japanese overseas development assistance is also scrutinised by the Japanese Board of Audit for regularity, efficiency and effectiveness. JICA has established a database of evaluation lessons which staff can search by keyword. Identifying such lessons when designing new projects has been mandatory since 2004 at the Ministry of Foreign Affairs (MoFA) and JICA. In terms of results-based management, JICA has adopted the cycle, "Plan, Do, Check, Act". Teams in JICA are now encouraged to base project design on the ultimate delivery of a service, rather than to see the successful completion of a project, as an end in itself. Results-based management is still in its early stages at the program level.

Denmark's DANIDA

Danida's performance management framework has three objectives: enhance the quality of development cooperation through a focus on results; to improve management and continuous learning through better information and monitoring; and to strengthen accountability through performance assessments and measurement.

New tools have been developed to further strengthen the performance framework, including guidelines for project management that apply to all programs exceeding DKK 5 million. They include, at the embassy level:

- Annual assessment of a country program: this self assessment includes reporting from program assessments done in connection with program reviews. It includes a qualitative review of i) the situation in the country; ii) general budget support; iii) program development, covering all sectors as well as cross-cutting issues and priority themes; and iv) reporting on Danida's Anti-corruption Action Plan
- Performance reviews of the bilateral development co-operation program are carried out every two years by the Quality Assurance Department: these peer reviews assess whether practices and activities agree with stated policies, goals, plans and procedures. They cover the internal organization and management system, the aid effectiveness agenda and financial management

In addition, at the program and project level, results-monitoring includes:

- Annual progress and financial reports on programs and projects
- Annual program reviews, which are mainly policy-oriented and focused on the overall implementation of national sector policies, strategies and programs, and on

the performance and relevance of Danish support

- Program and project completion reports finalized at the end of a program phase or project; and annual assessment by independent auditors of all Danish supported programs and projects, and a final audit at the end of all programs and projects

Asia Development Bank (AsDB)

The economic analysis of projects begins at project conception, is developed during project preparation, is monitored during project implementation, and is reviewed after project completion. This is done through a project completion report and, if appropriate, a project performance audit report to ensure that investment resources are used in a manner that is economic, efficient, and sustainable.

Economic analysis is carried out to assess the economic viability of a project in the context of a country's macroeconomic goals, performance, and outlook, and in the context of the goals, performance, and outlook of the relevant sector. All technically feasible alternative ways of achieving a project's objectives must be considered so that the least-cost technically feasible alternative can be identified. If this alternative is not selected, other economic principles must be applicable. If benefits cannot be valued, economic costs are assessed against project objectives to minimize the resources required to achieve the objectives. Where the outputs of a project can be quantified but not valued, economic efficiency can be assessed in terms of cost-efficiency alone.

Economic analysis should include the following elements:

- Review of macroeconomic context
- Review of sector context
- Demand analysis
- Identification of project rationale
- Identification of project alternatives
- Identification and comparison of project costs and benefits
- Assessment of project sustainability
- Distribution of project effects
- Sensitivity and risk analysis
- Identification of indicators for project performance monitoring system

International Commission for Aid Impact (ICAI)

ICAI's approach to assessing effectiveness and value for money is based on four guiding criteria built around the logical stages in the planning and delivery of aid programs.

1 Objectives: what is the program trying to achieve?

- 1.1 Does the program have clear, relevant and realistic objectives that focus on the desired impact?
- 1.2 Is there a clear and convincing plan, with evidence and assumptions, to show how the program will work?

- 1.3 Does the program complement the efforts of government and other aid providers and avoid duplication?
- 1.4 Are the program's objectives appropriate to the political, economic, social and environmental context?

2 Delivery: is the delivery chain designed and managed so as to be fit for purpose?

- 2.1 Is the choice of funding and delivery options appropriate?
- 2.2 Does program design and roll-out take into account the needs of the intended beneficiaries?
- 2.3 Is there good governance at all levels, with sound financial management and adequate steps being taken to avoid corruption?
- 2.4 Are resources being leveraged so as to work best with others and maximise impact?
- 2.5 Do managers ensure the efficiency and effectiveness of the delivery chain?
- 2.6 Is there a clear view of costs throughout the delivery chain?
- 2.7 Are risks to the achievement of the objectives identified and managed effectively?
- 2.8 Is the program delivering against its agreed objectives?
- 2.9 Are appropriate amendments to objectives made to take account of changing circumstances?

3 Impact: what is the impact on intended beneficiaries?

- 3.1 Is the program delivering clear, significant and timely benefits for the intended beneficiaries?
- 3.2 Is the program working holistically alongside other programs?
- 3.3 Is there a long-term and sustainable impact from the program?
- 3.4 Is there an appropriate exit strategy involving effective transfer of ownership of the program?
- 3.5 Is there transparency and accountability to intended beneficiaries, donors and UK taxpayers?

4 Learning: what works and what needs improvement?

- 4.1 Are there appropriate arrangements for monitoring inputs, processes, outputs, results and impact?
- 4.2 Is there evidence of innovation and use of global best practice?
- 4.3 Is there anything currently not being done in respect of the program that should be undertaken?
- 4.4 Have lessons about the objectives, design and delivery of the program been learned and shared effectively?

New Zealand's aid agency

Value for money is defined as:

“Achieving the best possible development outcomes over the life of an activity relative to the total cost of managing and resourcing that activity and ensuring that resources are used effectively, economically, and without waste.”

Assessing value for money requires that the overall benefit of an activity be weighed up and compared with the overall cost. This can often mean comparing and valuing outcomes (development impact and sustainability etc), with inputs (costs including time, effort, money expended and opportunity costs), a qualitative and quantitative exercise which ultimately requires judgement on the relative value of different criteria.

The three key aspects that need to be assessed are:

- Development outcomes
- Cost
- Effective and economic use of resources

A range of assessment tools are available to provide analysis and information to support good judgement in decision making. Assessment tools include:

- Comparisons with other activities
- Programme logic analysis
- Cost-effectiveness analysis
- Financial viability or cost-benefit analysis
- Economic cost-benefit analysis (CBA)
- Opportunity cost analysis
- Multi-criteria ranking
- New Zealand Aid Programme quality assurance controls

Assessment Tool	Description
Comparisons with other activities	Compare the costs of the activity, or the prices/rates charged for specific budget items with those in other activities with comparable outcomes and context. In some cases these may be available within MFAT or from other agencies. More specific guidance is also available or released from time to time, from New Zealand Government Agencies - specifically Treasury, State Services Commission and Office of Auditor General. This includes areas such as Capital Asset Management and Public Private Partnerships.
Programme logic analysis	A programme logic analysis assesses the internal logic of the programme and the fit between outcomes, outputs, inputs and budget.
Cost-effectiveness analysis	A cost-effectiveness analysis considers proposed outcomes and outputs, analyses the approach and budget and whether there are alternative ways of delivering the required outcome.
Financial viability or cost-benefit analysis	A financial analysis analyses the income and expenditure of the enterprise/service and the case for its viability.
Economic cost-benefit analysis (CBA)	Economic cost benefit analysis (CBA) assigns dollar values to all costs and benefits of an activity to produce a measure of the net monetary benefits (or cost) of the activity itself that can be compared or ranked against other activities. Commissioning a CBA depends on the scale and

Assessment Tool	Description
	<p>risk profile of the activity.</p> <p>More information on CBA is available from Finance or Sustainable Economic Development team members. Treasury also provides guidance to public sector agencies at www.treasury.govt.nz/publications/guidance/costbenefitanalysis</p>
Opportunity cost analysis	<p>An opportunity cost analysis is the value of what you could do with those resources if they were spent elsewhere. All expenditure has an opportunity cost. It is often useful to compare the activity under consideration, with other areas of the programme that will not receive funding if it goes ahead, i.e. to ascertain which has the more compelling business case.</p>
Procurement evaluation assessment	<p>MFAT commonly uses a procurement evaluation that involves multi-criteria ranking. The weighted-attribute model seeks to balance the trade-off between price and quality by weighting the criteria to reflect their relative importance. Each criterion in the proposal is scored, and each is multiplied by the relevant weighting to give a weighted score. The highest scoring offer is the preferred supplier. Such an approach is also possible in the non-commercial area, particularly when there are competing proposals.</p>
Quality assurance processes and tools	<p>Appraisal and peer review processes provide quality assurance on development effectiveness prior to final design. Monitoring and review provide reflection and evidence of progress in achieving development outcomes and revalidate value for money.</p>

Annex F: Tables

Table 1: Profile of CPEs Since 2008

Source: Information on evaluations completed drawn from DFATD evaluation reports website <http://www.acdi-DFATD.gc.ca/acdi-DFATD/acdi-DFATD.nsf/eng/NIC-316104532-LGZ>

Country of concentration	Evaluation report title	Year
Indonesia		Ongoing
Haiti		Ongoing
Afghanistan		Ongoing
Pakistan		Ongoing
West Bank/ Gaza		Ongoing
Ethiopia and Ghana	Countries being evaluated in a combined evaluation	Ongoing
Tanzania and Mozambique	Countries being evaluated in a combined evaluation	Ongoing
Caribbean Regional Program	Caribbean Regional Program Evaluation 2006-2011	2012
Peru	Evaluation of DFATD's Peru Program from 2005 to 2010	2012
Colombia	Colombia Country Program Evaluation 2006-2011	2011
Ukraine	Evaluation of DFATD's Ukraine Program from 2004 to 2009	2011
Honduras	Evaluation of DFATD's Honduras Program from 2002 to 2010	2011
Mali	Joint Evaluation of General Budget Support Operations in Mali 2003-2009	2011
Senegal	Evaluation of DFATD's Senegal Program from 2001 to 2010	2011
Vietnam	Evaluation of DFATD's Investments in Vietnam—July 2010	2010
Ethiopia	Ethiopia Country Program Evaluation 2003-2004 to 2008-2009	2010
Mozambique	Mozambique Country Program Evaluation—2004-2005 to 2008-2009	2010
Bangladesh	Bangladesh Country Program Evaluation 2003-2008	2009
Ghana	Evaluation of DFATD's Program in Ghana 1999-2005	2008
Bolivia	Bolivia Country Program Evaluation	2008

Country of concentration	Evaluation report title	Year
Cameroon	Canada—Cameroon Cooperation Program Evaluation	2008
Malawi	DFATD Malawi Country Program Evaluation-1998-2008	2010
Inter-American Program	Evaluation of DFATD's regional Inter-American Program from 2004-2005 to 2009-2010	2011

Table 2: Characteristics of Accountability Evaluations vs. Learning Evaluations

Characteristic	Accountability	Learning
Organizational context	Strong culture of accountability	Strong learning culture
Client	Senior management and external users (e.g. TB, Parliament)	Senior management and internal users (program managers) External users at country level
Coverage	Inclusive of all programming contexts (countries), in systematic fashion	Inclusive of all programming contexts (countries), in relation to need for information Possibility of focus on sectors/themes of relevance for Agency
Issues	Emphasis on strategic issues Emphasis on core issues for comparability across programs and over time	Emphasis on issues of importance to program planners Possibility of focusing on specific issues in specific contexts
Timing	Long time frame for evaluation to allow for best measurement of impacts Issue of timing not critical to evaluation	Time frame of evaluation adjusted to information needs of program units
Methodology	Emphasis on credibility of findings from perspective of external user Common methodology to facilitate comparisons across countries and over time	Emphasis on credibility and usefulness of findings from perspective of senior and program managers
Role of program staff	Independent from evaluation process, with exception of role in approving ToR	While not a must, could be engaged in, data collection, analysis and reporting for learning purposes
Structure of report	Speaks to wider audiences outside DFATD	Speaks to development professionals
Evaluation teams	Managed by unit external to program unit Real and perceived independence of teams from program staff	Allow for more engagement of program staff to ensure relevance to program information needs
Key stakeholders	Senior program managers Evaluation Committee	Senior program managers Program managers International and national partners

Table 3: Integration of ED Staff into the CPE Process

Stage of the CPE	Potential Role of DFATD ED Staff	Strengths	Weaknesses
Preparation of ToR	ED staff has main responsibility in consultation with Program	<ul style="list-style-type: none"> • Only practical possibility • Allows for Quality Assurance (QA) on RFP development 	<ul style="list-style-type: none"> • None
Preparation and management of the RFP and its assessment	ED staff has main responsibility in consultation with Program	<ul style="list-style-type: none"> • Necessary to allow for fair competition • Integrates with ED planning and budgeting 	<ul style="list-style-type: none"> • None •
Inception/Planning Mission - Maximum ED Involvement	ED staff participate alone	<ul style="list-style-type: none"> • Allows ED maximum influence in operational planning of CPE • Allows ED staff to build relationship with potential local consultants and with DFATD staff in-country 	<ul style="list-style-type: none"> • Denies external evaluation team members direct knowledge of data availability and logistics issues • Implies DFATD ED staff have greater responsibility/ accountability for operational plan for CPE (and results) and thus limited ability to undertake QA
Inception Planning Mission - Shared ED Involvement	ED staff participate jointly with consultant team leader	<ul style="list-style-type: none"> • Allows ED staff to influence operational plan • Allows ED staff and operational team lead to gain joint understanding of local capacity • Gives consultant team lead access to data availability and logistics information 	<ul style="list-style-type: none"> • May limit ED staff sense of ownership of operational planning for the CPE • Is more expensive
Primary Data Collection Mission - Maximum ED Involvement	ED staff participate in data collection and preliminary (in-country reporting)	<ul style="list-style-type: none"> • Gives ED staff direct knowledge of evidence base. • Allows ED staff to participate in and understand development of preliminary findings and conclusions • Increases confidence of ED staff in CPE findings. 	<ul style="list-style-type: none"> • May cause some interviewees/key informants to be less candid. • Raises issues of team leadership and may cause pushback from local consultants. • Raises question of what to do if DFATD staff do

Stage of the CPE	Potential Role of DFATD ED Staff	Strengths	Weaknesses
			not deliver on commitments. <ul style="list-style-type: none"> • Compromises ED staff QA role.
Primary Data Collection - Reduced ED Involvement	ED staff do not take part in data collection in the field but do join field team for preparation and presentation of preliminary findings to DFATD and other stakeholders in the country.	<ul style="list-style-type: none"> • May encourage key informants to be more candid. • Preserves lines of authority in the evaluation team • Allows DFATD ED staff to observe and better understand how evidence base is compiled and linked to preliminary findings • Maintains ED staff independence for QA 	<ul style="list-style-type: none"> • ED staff presence in workshops to develop preliminary findings may inhibit full debate in evaluation team. • ED staff may still have limited understanding of the strengths and weaknesses of the data collection methods and processes.
Primary Data Collection - Minimal ED Involvement	Primary data collection mission and preliminary reporting by consultants only.	<ul style="list-style-type: none"> • Allows for maximum perceived independence by evaluation team. • Clear lines of authority and responsibility within the evaluation team • Interviewees and key informants may be more candid • Maintains ED staff independence for QA 	<ul style="list-style-type: none"> • Limits ED staff ability to managed the data collection process • May weaken ED staff understanding of the evidence base and the strength of findings, conclusions they will have to present to EC.

Annex G: List of interview respondents

DFATD-Development

Executive and Senior Program Managers

Greta Bossenmaier	Senior Vice-President
Darren Schemmer	Vice-President, Partnership Branch
David Morrison	VP, Geographic Programs
Lise Filiatrault	Regional Director General
Bob Johnston	Regional Director General, Eastern Europe
Susan Steffen	A/Regional Director General, West and Central Africa
Jeff Nankivell	Regional Director General, Asia
Isabelle Bérard/ Sébastien Séguin	Director General, Planning, Operations and Specialists Directorate/ Director, Strategic Planning and Operations Division

Program managers

Andres Scyner	
Antoine Chevrier	
Gina Watson	Country Program Manager, Ukraine
Joshua Tabbeh	
Kate Stefanuck	International Development Program Manager, Honduras
Louis Verret	Development Officer, Afghanistan
Luc Pincince	
Louise Clément	

Evaluation Division

Michelle Samné	Evaluation Manager
Joelle Barbot	Evaluation Manager
Pierre Tremblay	Team leader
Frantz Pierre-Jerome	Evaluation Manager

Other DFATD-Development

Goberdhan Singh	Former, Director General, Evaluation
Marc André Fredette	Former, Director responsible for corporate monitoring Tools

Jean-Marc Lafrenière/ Steve Kirk

Audit Managers

Evaluation Committee

Maureen O'Neill

Committee member

James McDavid

Committee member

David Runnalls

Committee member

Richard Carey

Committee member

Emile Gauvreau

Committee member

Niels Dabelstein

Committee member

Consultants

English Group 1

Charles Lusthaus and Gerry Cooney, Universalialia

Margot Rothman, interalia

Allan Barry and Murray Smith, IDMAG

Susanne Duska and Keith Ogilvie) DADA International Inc.

English Group 2

Werner Meier, RBMG

Krista Ladouceur, Cowater

French group

Alain Boisvert and Gilbert... Inter Alia

Yvan Bergeron, Econotec

Elisabetta Mocaró and R. Boers, Universalialia

Marie-Hélène Adrien

Universalialia

Philippe Bacle

Baastel

Other donors

Cynthia Clapp-Wincek

Director Office of Learning, Evaluation and Research, USAID

Ian Murray/Elizabeth Robin

Evaluation Managers, DFID