DFID **Department for International Development**

PLANNING EVALUABILITY ASSESSMENTS

A SYNTHESIS OF THE LITERATURE WITH RECOMMENDATIONS

Report of a study commissioned by the Department for International Development

Planning Evaluability Assessments

A Synthesis of the Literature with Recommendations

By Dr Rick Davies

Cambridge, August 2013

Table of Contents

ANNEXES

Abbreviations

AusAID	Australian Agency for International Development
CDA	The Collaborative for Development Action (USA)
DFID	Department for International Development (U.K)
EBRD	European Bank for Reconstruction and Development
EC	European Commission
IADB	Inter–American Development Bank
IDRC	International Development Research Centre (Canada)
Intervention	Used as a synonym for project
ILO	International Labour Organisation
NDC	Netherlands Development Cooperation
MoD	Ministry of Defence
NORAD	Norwegian Agency for Development Co-Operation
OECD DAC	Organisation for Economic Cooperation and Development – Development Assistance Committee
Project	A time bound collaborative enterprise planned and designed to achieve an objective, with a dedicated budget
Program/me	Used as a synonym for project
SIDA	Swedish International Development Cooperation Agency
ToC	Theory of Change
ToR	Terms of Reference
UNIFEM	United Nations Fund for Women
UNODC	United Nations Office on Drugs and Crime
USAID	United States Agency for International Development

Acknowledgements

Acknowledgements

EXECUTIVE SUMMARY

S1.   The purpose of this synthesis paper is to produce a short practically oriented report that summarises the literature on Evaluability Assessments, and highlights the main issues for consideration in planning an Evaluability Assessment. The paper was commissioned by the Evaluation Department of the UK Department for International Development (DFID) but intended for use both within and beyond DFID.

S2.   The synthesis process began with an online literature search, carried out in November 2012. The search generated a bibliography of 133 documents including journal articles, books, reports and web pages, published from 1979 onwards. Approximately half (44%) of the documents were produced by international development agencies. The main focus of the synthesis is on the experience of international agencies and on recommendations relevant to their field of work.

S3.   Amongst those agencies the following OECD DAC definition of evaluability is widely accepted and has been applied within this report: *"The extent to which an activity or project can be evaluated in a reliable and credible fashion"*.

S4.   Eighteen recommendations about the use of Evaluability Assessments are presented here, based on the synthesis of the literature in the main body of the report. The report is supported by annexes, which include an outline structure for Terms of Reference for an Evaluability Assessment.

PURPOSE

S5.   An Evaluability Assessment should examine evaluability: (a) in principle, given the nature of the project design, and (b) in practice, given data availability to carry out an evaluation and the systems able to provide it. In addition it should examine the likely usefulness of an evaluation. Results of an Evaluability Assessment should have consequences: for the design of an evaluation, the design of an M&E Framework, or the design of the project itself. An Evaluability Assessment should not be confused with an evaluation (which should deliver the evaluative judgements about project achievements). (See page 7)

S6.   Many problems of evaluability have their origins in weak project design. Some of these can be addressed by engagement of evaluators at the design stage, through evaluability checks or otherwise. However project design problems are also likely to emerge during implementation, for multiple reasons. An Evaluability Assessment during implementation should include attention to project design and it should be recognised that this may lead to a necessary re-working of the intervention logic. (See page 8)

PLANNING

S7.   Evaluability Assessments do not need to be limited to specific projects, although that is their most common focus. They can also be applied to portfolios of activities, legislation and other policy initiatives, country and sector strategies and partnerships which may have longer time frames. (See page 10)

S8.   The timing of an Evaluability Assessment will depend on the expected outcomes of the assessment: to improve the project design prior to approval; or to inform the design of an M&E framework in the inception period; or to decide if an evaluation should take place later on; or to inform the specific design of an evaluation that has now been planned for. Early assessments may have wider effects on long term evaluability but later assessments may provide the most up to date assessment of evaluability. (See page 11)

S9.   Locally commissioned Evaluability Assessments are likely to have the most support and generate the most value. However, other complimentary strategies may be useful, including centrally provided technical advice, screening of a random sample of projects in areas where little assessment work has been done to date and mandatory assessments for projects with budgets above a designated size. (See page 12)

S10.   Ideally Evaluability Assessments would be carried out by independent third parties, not project managers or those commissioned to carry out a subsequent evaluation. Where Evaluability Assessments are carried out by an independent third party they can examine the feasibility of alternative evaluation designs, but they should not specify the designs to be used by an evaluation team. (See page 13)

S11.   While recognising the vast variation in project designs and sizes, past Evaluability Assessment practice suggests two time budgets should be considered: Five days for desk-based studies with no country visits, and up to two weeks for in-country assessments (both per project). (See page 14)

S12.   Evaluability Assessments can be carried out at a small fraction of the cost of most evaluations. They can offer good value for money, if they are able to influence the timing and design of subsequent evaluations. (See page 15)

PROCESS

S13.   No specific stage model can be recommended for an Evaluability Assessment from amongst those that exist. However, common steps include: (a) Identification of project boundaries and expected outputs of the Evaluability Assessment, (b) Identification of resources available for the assessment, (c) Review of the available documentation, (d) Engagement with stakeholders, (e) Development of recommendations, (f) Feedback findings to stakeholders. Recommendations should cover: (i) Project logic and design,

(ii) M&E systems and capacity, (iii) Evaluation questions of concern to stakeholders, (iv) Possible evaluation designs. (See page 17)

S14.   An examination of guidance documents produced by eight international agencies suggests that an Evaluability Assessment should attend to three broad types of issues:

- The program design
- The availability of information
- The institutional context

These relate closely to the three purposes of Evaluability Assessment discussed above.

For each of these main issues a number of specific criteria and associated questions will be relevant. These are summarised in three tables on pages 20–23.

The division of attention across these areas will be subject to the timing of an Evaluability Assessment, with design being the main focus at a quality assessment stage and information availability and conduciveness becoming relatively more important during implementation and immediately prior to an evaluation.

S15.   Evaluability Assessment checklists should be used. They encourage comprehensive coverage of relevant issues, and visibility of those that are not covered. They can be used as stand-alone tools along with ratings, or be supported by comment and analysis or have a more background role informing the coverage of a detailed narrative report. This report provides a three part checklist of issues to be addressed. (See pages 20–23)

S16.   The aggregation of individual judgements within an Evaluability Assessment into a total score is good practice because it enables comparisons of evaluability across projects and across time, and thus lessons learned from differences in evaluability. The use of minimum threshold scores is not advisable unless there are very goods grounds for defining such a threshold. (See page 26)

S17.   Where scored checklists are used there should be explicit weightings, to avoid mistaken assumptions about all criteria being equally important. Weightings can be either built in by the checklist designer or provided by the checklist user – as part of their assessment. If possible, explanations should be sought for weightings, in order to make judgements more transparent. (See page 27)

S18.   The results generated by a scored checklist should be seen as an index of difficulty, which then needs to be responded to by program managers and/or evaluators when they are commissioning or planning an evaluation. Not as a final judgement on evaluability, given the range of evaluation methods and purposes that exists. (See page 28)

S19.   Outputs of an Evaluability Assessment should include both assessments and recommendations. Assessments should cover: (a) evaluability of the project, referring

both to the project design and the availability of information, and (b) the practicality and utility of an evaluation. Recommendations can refer to: (a) changes in project design and associated M&E systems to make it more evaluable, (b) options for evaluation timing, evaluation questions and evaluation methods, to help ensure the usefulness of an evaluation. (See page 30)

Recommendations should inform the design of Terms of Reference (ToR) for an evaluation, but not pre-empt the design of an evaluation.

Annex G provides an outline structure for Terms of Reference for an Evaluability Assessment.

CONCERNS

S20.   While there is limited systematic evidence on the effectiveness of Evaluability Assessments the relatively low costs of Evaluability Assessments means that they only need to make modest improvements to an evaluation before their costs can be recovered. (See page 32)

S21.   The biggest risk of failure facing an Evaluability Assessment is likely to be excessive breadth of ambition: reaching into evaluation design or evaluation itself. This risk may be higher when Evaluability Assessments are undertaken in-country in association with stakeholders, versus at a distance via a desk based analysis. It should also be recognised that Evaluability Assessment may be seen as challenging, if there are already some doubts about a project design. (See page 33)

SUPPORTING INFORMATION

S22.   The main report is supported by six annexes, including the methodology used for the literature review, sources of other example checklists in addition to the checklists proposed on pages 20-23 of the report and an outline structure for the Terms of Reference for an Evaluability Assessment. The complete bibliography, including abstracts and hypertext links, is now available online at http://mande.co.uk/blog/wp-content/uploads/2013/02/Zotero-report.htm.

S23.   There has been a resurgence in the use of Evaluability Assessment but not yet in the published literature on Evaluability Assessment. Guidance material is becoming more available but reviews of the use of Evaluability Assessments are still scarce. The online bibliography produced as a part of this synthesis report should be periodically updated and publicised to make sure future experiences with Evaluability Assessment are widely accessible, and open to further reviews. (See page 34)

1   INTRODUCTION

1.1   The purpose of this synthesis paper is to produce a short practically oriented report that summarises the literature on Evaluability Assessments, and highlights the main issues for consideration in commissioning an Evaluability Assessment. The paper was commissioned by the Evaluation Department of the UK Department for International Development (DFID) but intended for use both within and beyond DFID. See Annex C for the Terms of Reference.

1.1   What experience is there to learn from?

1.2   The synthesis process began with a literature search, the details of which are given in Annex B. The search generated a bibliography of 133 documents including journal articles, books, reports and web pages, covering the period 1979 to 2012. Of these 59% described actual examples of Evaluability Assessments, 13% reviewed experiences of multiple kinds of Evaluability Assessments, 28% were expositions on Evaluability Assessments, with some references to examples, 10% were official guidance documents on how to do Evaluability Assessments and 12% were Terms of Reference for Evaluability Assessments. Almost half (44%) of the documents were produced by international development agencies[1].The majority of the remaining documents were produced by state and national agencies in the United States.

1.3   Very few of the Evaluability Assessments carried out by international agencies make any reference to prior experiences with Evaluability Assessments. A few made reference to the more widely cited American commentators in the field, such as Wholey, Thurston, Smith, Leviton and Trevisan[2]. An early output of this synthesis study has been an online bibliography of documents on Evaluability Assessment, many of which include hypertext links to the documents themselves.

http://mande.co.uk/blog/wp-content/uploads/2013/02/Zotero-report.htm

This bibliography has since been publicised via M&E email lists and websites, which should help widen exposure to the range of Evaluability Assessment practice and experience that exists.

1.4   Expanding access to past experience is timely. The number of Evaluability Assessments, and reports and papers about them, appears to have grown substantially in the last five years, as the chart on the next page shows[3].

1.5   Within development aid agencies interest in Evaluability Assessment appears to be growing. In the last twelve month guidance on Evaluability Assessment have been

---

[1] These are descriptive rather than inferential statistics, describing what was found, and not necessarily what exist in total, if there was a much more extensive search. The figures add up to more than 100% because some documents belonged to more than one category.

[2] Whose publications represented 14% of all the documents found. See Annex D for a description of their stage views of the Evaluability Assessment process, along with those of international agencies

[3] Caveats concerning this data are noted in Annex B.

developed by ILO, CDA, IDRC, ERBRD and UNODC[4]. In 2012 the DFID Evaluation Department has funded 12 Evaluability Assessments requested by its country offices. AusAID Indonesia commissioned 4 Evaluability Assessments during the same period.
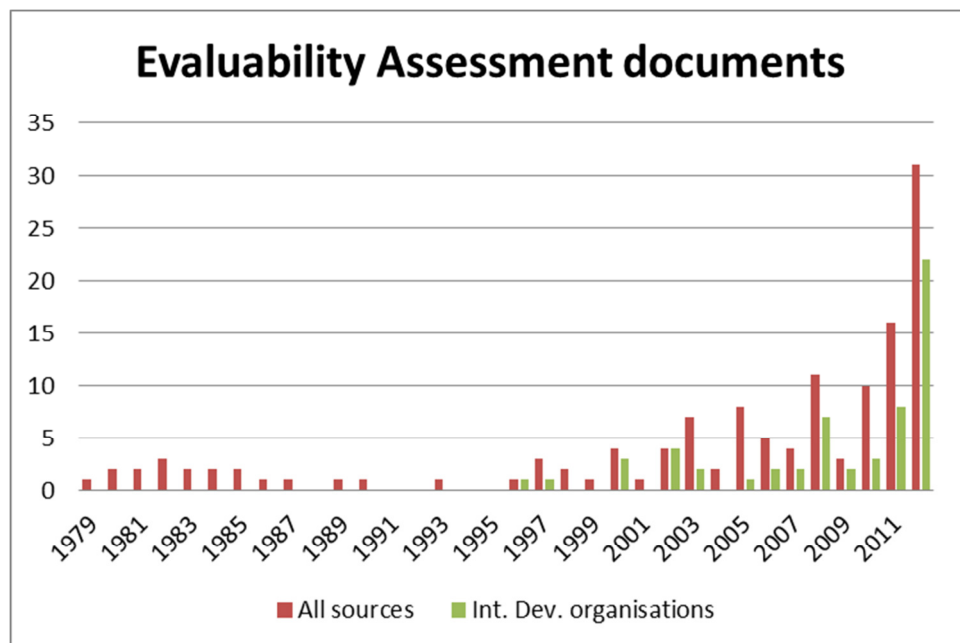
---

2   PURPOSE

2.1   What is evaluability?

2.1   Amongst international development agencies there appears to be widespread agreement on the meaning of the term. This OECD DAC definition is widely quoted and used[5]:

"The extent to which an activity or project can be evaluated in a reliable and credible fashion"

2.2   What is an Evaluability Assessment?

2.2   Descriptions of what constitutes an Evaluability Assessment are more elaborate and varied. The concept of evaluability is often used in two different but complimentary ways. One is "in principle" evaluability, which looks at the nature of a project design, including its Theory of Change (ToC) and asks if it is possible to evaluate it as it is described at present. The second is "in practice" evaluability and looks at the availability of relevant data, as well as systems and capacities which make that data available.

2.3   In addition, most Evaluability Assessments extend their interests beyond evaluability itself. The most common extension is an inquiry into the practicality and usefulness of doing an evaluation through discussions with stakeholders (e.g. as used by UNIFEM, AusAID, EC, NDC, and EBRD). Other extensions of purpose focus on specific uses of Evaluability Assessment findings to inform the design of an expected evaluation, or more generally, the project's overall M&E framework. Improvements in the project design itself is usually a less explicit purpose, but can be an unavoidable consequence of some Evaluability Assessment findings, such as lack of clarity about expected causal linkages between expected outputs and outcomes.

2.4   Evaluability Assessments can overlap in purpose with other activities. They can segue into mini-evaluations, especially in the eyes of stakeholders being contacted. Some writers like Leviton (2010) have gone so far as to argue that Evaluability Assessment could be called "exploratory evaluation". This seems unhelpful and more likely to cause confusion and loss of focus. Evaluability Assessments can also overlap with quality assurance processes focusing on project design (e.g. UNIFEM 2012). After a series of independent Evaluability Assessments over nearly a decade the IADB has recently sought to integrate evaluability assessment into its design quality assurance procedures, albeit backed up by an independent audit (Office of Evaluation and Oversight 2011). While this seems to be a positive development it is unlikely to be sufficient for many organisations, given that many evaluability issues may not become visible until project implementation begins.

2.5   UNIFEM have usefully commented that "It is important to note that Evaluability Assessment does not replace good programme design and monitoring functions; rather,

---

[5] Including IFAD, UNODC, OECD, SIDA, ILO, DFID, NORAD and NDC

it is a tool that helps managers to verify whether these elements are in place and to fill any common gaps.”

Recommendation 1: An Evaluability Assessment should examine evaluability: (a) in principle, given the nature of the project design, and (b) in practice, given data availability to carry out an evaluation and the systems able to provide it. In addition it should examine the likely usefulness of an evaluation. Results of an Evaluability Assessment should have consequences: for the design of an evaluation, the design of an M&E Framework, or the design of the project itself. An Evaluability Assessment should not be confused with an evaluation (which should deliver the evaluative judgements about project achievements).

2.3 Why are Evaluability Assessments needed?

2.6 From Wholey in the 1970's onwards it appears that the main concern of writers on Evaluability Assessment has been with the number of poor quality evaluations that are being produced. Associated with this has been concern about the cost of those evaluations and the need for some economy of effort (Leviton, 2010, Ogilvie et al. 2011).

2.7 Underlying the problem of poor quality evaluations is the problem of poor quality project designs. Reviewing Wholey's findings on evaluation in the 1970's Dawkins (2010) notes that:

Many studies found null or negative results due to:
- Programs not fully implemented or did not exist
- Goals were “grant goals”
- Lack of logic in design
- Lack of use due to lack of “ownership” or agreement with the focus of the results

Many program goals and objectives exist only on paper;
- Or, they were never articulated
- Or, stakeholders disagree about them
- Or, program reality is not consistent with them

2.8 Recently Ruben (2012) has repeated this argument, more forcefully. Reflecting on the Netherlands development aid program context he noted a “Growing number of pseudo evaluations” and in regard to private sector programs that “Two thirds of executed ‘evaluations’ cannot be used”. The reasons why include:
- evaluation agency not fully independent
- stated objectives too broad/vague
- no clear indicators defined
- data at too aggregate level
- absence of baseline data
- no representative sampling
- too general intervention theory

2.9   At least five of the above problems noted by Ruben have their roots in the design process. The IADB's use of Evaluability Assessments over the last decade has been oriented towards addressing such design problems. AusAID Indonesia's more recent experience with Evaluability Assessments has also been oriented in this direction, taking place after projects have been approved, but well before any evaluations have been scheduled. DFID's involvement of evaluation expertise in the development of Business Cases for new projects seems to be intended to serve the same purpose.

2.10   In an ideal world projects would be well designed. One aspects of their good design would be their evaluability. Evaluability Assessments would not be needed, other than as an aspect of a quality assurance process closely associated with project approval (e.g. as used by IADB). In reality there are many reasons why approved project designs are incomplete and flawed, including:

- Political needs may drive the advocacy of particular projects and override technical concerns about coherence and quality.
- Project design processes can take much longer than expected, and then come under pressure to be completed.
- In projects with multiple partners and decentralised decision making a de facto blueprint planning process may not be appropriate. Project objectives and strategies may have to be "discovered" through on-going discussions.
- Expectations about how projects should be evaluated are expanding, along with the knowledge required to address those expectations.

2.11   In these contexts Evaluability Assessments are always likely to be needed in a post–project design period, and will be needed to inform good evaluation planning.

Recommendation 2: Many problems of evaluability have their origins in weak project design. Some of these can be addressed by engagement of evaluators at the design stage, through evaluability checks or otherwise. However project design problems are also likely to emerge during implementation, for multiple reasons. An Evaluability Assessment during implementation should include attention to project design and it should be recognised that this may lead to a necessary re-working of the intervention logic.

3   PLANNING

3.1   What kinds of activities can be assessed?

3.1   Evaluability Assessments are typically focused on individual projects and their evaluability. Approximately 60% of Evaluability Assessments listed in the bibliography are in this category. However their ambit has expanded over time. In the field of development aid they have also included:

- Sets of projects of a kind. Such as Sida's funding of 28 democracy and human rights projects in Latin America and South Africa (Poate, 2000)
- Policy areas, where the total number of relevant projects may not yet be known. Such as DFID's work on empowerment and accountability and the DFID Strategic Vision for Girls and Women (Davies, et al, 2012)
- Country strategies. Such as the UNEG's Evaluability Assessments of the Programme Country Pilots Delivering as One UN (UNEG, 2008)
- Strategic plans. Such as the Evaluability Assessment of the UNIFEM Strategic Plan (2008-2013) (IOD/PARC, 2011)
- Work Plans. Such as the UNDP's Evaluability Assessment of UN Women Pacific Sub Regional Office Annual Work Plan and Programme Plans (UNDP, 2007)
- Partnerships. Such as NORAD's evaluability study of partnership initiatives supporting Millennium Development Goals 4 & 5 (Plowman et al, 2011)

3.2   Elsewhere, Evaluability Assessments have also been carried out on:

- The implementation of legislation (Jung, 1980)
- The introduction of information technologies (LMIT–BPS, 2008)

> Recommendation 3: Evaluability Assessments do not need to be limited to specific projects, although that is their most common focus. They can also be applied to portfolios of activities, legislation and other policy initiatives, country and sector strategies and partnerships.

3.2   When to carry out an Evaluability Assessment?

3.3   Different agencies have used Evaluability Assessments at different points in the project management cycle.

- **At the project design stage**: The IADB uses Evaluability Assessments as part of the project design approval process. They take place before the projects have been approved. The EC (Evalsed, 2009) and UNODC (Gunnarsson, 2012) also propose their use at this stage.
- **At the M&E Framework stage**: AusAID Indonesia use Evaluability Assessments after projects have been approved but prior to or during the development of an M&E Plan for the project. DFID has also increased its usage of Evaluability Assessments during the inception period of project implementation.
- **Prior to evaluations**: DFID, NORAD, SIDA and others have used Evaluability Assessment after projects have been in operation for some time, and before they are evaluated.

- **During evaluations**: USAID and other agencies have incorporated Evaluability Assessments as a stage in the evaluation process, prior to evaluation design (Dunn 2008). In these circumstances it is in effect assumed that an evaluation will be possible, but it will need to be informed by evaluability constraints.

3.4   Monk (2012) has documented the timing of Evaluability Assessments as used by 13 international organisations, differentiating between: (a) Use at the beginning of the project – by 5 organisations, (b) Use just before the evaluation – by 9 organisations (See Annex E).

3.5   The EBRD has argued strongly for early use of Evaluability Assessment: "definitions state or imply that an Evaluability Assessment is something carried out before the conduct of an ex–post evaluation. While this type of assessment would be useful to help the Evaluation department avoid wasting time and effort trying to evaluate something not capable of being evaluated in a reliable and credible way, it is too late to do anything to change the reality" (Leonard and Eulenberg. 2012).

<u>Recommendation 4</u>: The timing of an Evaluability Assessment will depend on the expected outcomes of the assessment: to improve the project design prior to approval; or to inform the design of an M&E framework in the inception period; or to decide if an evaluation should take place later on; or to inform the specific design of an evaluation that has now been planned for. Early assessments may have wider effects on long term evaluability but later assessments may provide the most up to date assessment of evaluability.

3.3   Mandatory or voluntary?

3.6   IADB Evaluability Assessments are compulsory in the sense that they are carried out on a random sample basis, by people other than those responsible for the management of the sampled projects. This process is managed by the Office of Evaluation and Oversight. Random sampling means there can be quality control over a large number of projects, despite limited resources.

3.7   A recent discussion paper by the EBRD has argued that "Evaluability Assessments should become a routine part of the approval process for new EBRD operations with a minimum acceptable level of evaluability established. It is suggested that this start with grants (technical cooperation and so on) with a progressive roll – out to other operations" (Leonard et al 2012).

3.8   DFID Evaluability Assessments are voluntary and initiated by the persons responsible for the projects that will be assessed. The Evaluation Department provides Evaluability Assessments to project managers, on request, using external consultants available on a call down basis. The main incentive for their continued and wider use is the immediate value they are seen to provide. Most noticeably by improving the design of Terms of Reference for evaluations.

3.9   AusAID Indonesia Evaluability Assessments appear to be initiated by the country program's Performance and Quality Unit, on an as-needed basis.

<div style="background-color: yellow; border: 1px solid black; padding: 10px;">

<u>Recommendation 5</u>: Locally commissioned Evaluability Assessments are likely to have the most support and generate the most value. However, other complimentary strategies may be useful, including more centrally provided technical advice, screening of a random sample of projects in areas where little assessment work has been done to date and mandatory assessments for projects with budgets above a designated size[6].

</div>

3.4   Who should do it?

3.10   **Inside or outside?** Evaluability Assessments can be carried out by staff within an organisation that is implementing or funding a project or by others outside who are contracted. IADB has used its own staff, from within the Office of Evaluation and Oversight. DFID and USAID have contracted outside parties. IADB's use of its own staff is possible because of the limited scope of the task, being based on deskwork only. However, DFID have contracted out Evaluability Assessments that range in size from small to large scale (1 week to months). Externally contracted Evaluability Assessments are the most common practice amongst the examples found during this review.

3.11   **Kinds of expertise**: USAID experience suggests that a mix of evaluation and subject matter expertise is desirable. Evaluation expertise is necessary to address methodological issues around data and its analysis but subject matter expertise is needed to assess the plausibility of the expected effects of interventions, the quality of evidence and potential usefulness of findings. In short desk based assessments this mix of expertise may not be possible, but in longer field based assessments, involving stakeholder consultations, it should be.

3.12   **Separate or joint contracts?** Any process of planning an evaluation necessarily involves some form of Evaluability Assessment, such as checking the status of the project's Theory of Change and the availability of relevant data. Some evaluation reports have separate sections specifically on evaluability. Other evaluations are preceded by a separate evaluability study, by those who will subsequently do the evaluation. In other cases the Evaluability Assessment will be done by an independent third party who will not undertake the evaluation.

3.13   If there is significant initial doubt as to the value of doing an evaluation then a separately contracted Evaluability Assessment would seem best. This would minimise a possible conflict of interest i.e. the contractor would not be inclined to downplay the difficulties in order to avoid losing an evaluation contract. One example was found where a company contracted to do an evaluation did conclude through a prior Evaluability Assessment that the planned evaluation would not be feasible (Snodgrass, Magill, and Chartock 2006). However the company concerned had an encompassing

---

[6] Possibly a high level to begin with, to limit additional workloads and help build up experience

large scale contract for the evaluation for a range of projects and would not have been disadvantaged by being so forthright.

3.14   Complications are likely to arise when the scope of an Evaluability Assessment extends into the design of an evaluation, if the evaluation is expected to be implemented by another party. The second party may not fully agree or understand the rationale behind the design. One USAID guidance document has sensibly limited Evaluability Assessments to the examination of the feasibility of alternative designs, not the choice of specific designs. Evaluability Assessments that are carried out as a stage within an evaluation, all managed by the same team, do not have to deal with potentially conflicting design requirements.

<u>Recommendation 6</u>: Ideally Evaluability Assessments would be carried out by independent third parties, not project managers or those commissioned to carry out a subsequent evaluation. Where Evaluability Assessments are carried out by an independent third party they can examine the feasibility of alternative evaluation designs, but they should not specify the designs to be used by an evaluation team.

Applicable to evaluation contracts?

"A 'conspiracy of optimism' exists between MoD and industry, each having a propensity, in many cases knowingly, to strike agreements that are so optimistic as to be unsustainable in terms of cost, timescale or performance"

http://www.rusi.org/downloads/assets/Acquisition_Focus_0207.pdf.

3.5   How long does it take to do an Evaluability Assessment?

3.15   In the earliest experiences with Evaluability Assessments in America, in the 1970's they take anywhere between two weeks and a year (Ruteman cited by Monk, 2012). Amongst the 29 examples of Evaluability Assessments by international development agencies the duration ranges from two days to four months, with one week being perhaps the most common[7]. The quickest Evaluability Assessments were typically desk based exercises, utilising readily available documents, notably those by IADB, which take an estimated two days. DFID's desk-based Evaluability Assessments have taken five days each. These seem to be the most common type of Evaluability Assessment undertaken by DFID in recent times.

3.16   AusAID Indonesia's recent set of Evaluability Assessments, carried out in–country, have taken around two weeks. Some USAID Evaluability Assessments which have focused on one project per country have also spent two weeks in-country.

3.17   The longest Evaluability Assessments involved the assessment of multiple projects, and may involve country visits and consultations with stakeholders (NORAD, SIDA, and DFID). The NORAD study took 24 weeks and involved a core team of two

---

[7] Note however that this information was missing from almost half the documents

consultants doing a desk-based analysis of five country partnerships. The SIDA study, which covered 30 projects in 4 countries over a four months period, took an average of two days field work per project, and was undertaken jointly by an international and local consultant. Desk-based Evaluability Assessment of large portfolios of projects may also take some months e.g. of DFID's Empowerment and Accountability portfolio of projects (Davies et al. 2012). This category of Evaluability Assessment is the least common.

3.18   Reflecting on the SIDA experience Poate et al (2000) noted "It is estimated that an Evaluability Assessment needs an average of two to three days per project dedicated to desk-based review of documentation and four to five days devoted to fieldwork. One day per stakeholder group would be necessary with the right conditions for holding a workshop-style event (with an appropriate environment, materials, etc.). As there are usually at least three main stakeholder groups, this would cover at least four days including preparation and write-up. A fifth day would be necessary to present the results to a selection of the different stakeholders in one event where the differences emerging could be presented, confirmed and commented upon".

3.19   For in-country Evaluability Assessments, the biggest influence on time requirements is the need for stakeholder consultations[8]. These will be prioritised when the utility function of an Evaluability Assessment is being emphasised. Consultations are more likely to have a higher priority later in the project cycle when a project is well underway, rather than at the design or early inception stage.

> Recommendation 7: While recognising the vast variation in project designs and sizes, past Evaluability Assessment practice suggests at least two time budgets should be considered: Five days for desk-based studies with no country visits, and up to two weeks for in-country assessments (both per project).

3.6   What does an Evaluability Assessment cost?

3.20   Not surprisingly, cost information is not readily available in most documents. The following costed examples have been found:
- Evaluability Assessments commissioned by the DFID Evaluation Department, costing an average of £4000. These involved desk-based work only.
- Evaluability Assessments commissioned by AusAID Indonesia, costing an average of A$8,000 (fees only, excluding costs). These involved field work in country.
- An Evaluability Assessment for UNICEF Tanzania in 2008, which cost US$14,766. This involved two week's field work in-country.

Costs of Evaluability Assessments involving multiple in-country visits such as those funded by NORAD, SIDA and DFID have not been identified but are likely to be very much higher.

3.21   From the information that has been found it seems that most Evaluability Assessments will cost a small fraction of the cost of an evaluation. In as much as they help avoid

---

[8] To address issues about the likely utility of an evaluation, but also of the evaluability of the project design, as detailed on pages 20-23

unproductive evaluations and improve the quality of evaluations that are carried out, they offer good value for money.

3.22   Given current pro-transparency policies of most large international development agencies, it could be expected that the costs of Evaluability Assessment would be routinely disclosed within the text of the reports themselves, as exemplified by a UNICEF Evaluability Assessment in Tanzania (Yantio, 2008).

> Recommendation 8: Evaluability Assessments can be carried out at a small fraction of the cost of most evaluations. They can offer good value for money, if they able to influence the timing and design of subsequent evaluations.

4   PROCESS

4.1   What process should be followed?

4.1   In the literature on Evaluability Assessment there are many different views of how the Evaluability Assessment should work. Twelve examples are show in Annex D. They include a mixture of sequences of activities and checklists of activities.

4.2   The outline below is an interpretative synthesis of their contents. It may not be representative of the typical Evaluability Assessment.

1. Define the boundaries of the project
   - Time period, geographical extent, and relevant stakeholders
   - Agree on expected outputs of the Evaluability Assessment

2. Identify the resources available
   - Documents
   - Stakeholders

3. Identify and review documents, including
   - The program logic/theory of change/results chain
   - Its clarity, plausibility, ownership
     - Information systems
     - Availability, relevance and quality of data, capacity of systems and staff to deliver what is needed
   - Examine implementation relative to plans

4. Engage with stakeholders
   - Identify their understandings of program purpose, design and implementation, including areas of agreement and disagreement
   - Identify their expectations of an evaluation, it objectives, process and use
   - Clarify and fill in gaps found in document review

5. Develop conclusions and make recommendations, re:
   - Project logic improvements
   - M&E systems and capacity development
   - Evaluation questions of priority interest to stakeholders
   - Possible evaluation designs

6. Feedback findings and conclusions to stakeholders

4.3   Desk based Evaluability Assessments will have little opportunity for stakeholder engagement, whereas in–country Evaluability Assessments will be able to give this much higher priority.

4.4   Leviton (2010) has criticised stage models of Evaluability Assessments as being unrealistic, in that in most situations the process is more iterative than linear. In reality a review of documents will lead to stakeholders but contact with stakeholders will also

lead to documents. Meetings with stakeholders can be difficult to organise, and in practice are used more opportunistically. If there are repeat meetings they will be used to gap fill, rather than proceed to the next step in an idealised process.

> Recommendation 9: No specific stage model can be recommended for an Evaluability Assessment from amongst those that exist. However, common steps include: (a) Identification of project boundaries and expected outputs of the Evaluability Assessment, (b) Identification of resources available for the assessment, (c) Review of the available documentation, (d) Engagement with stakeholders, (e) Development of recommendations, (f) Feedback findings to stakeholders. Recommendations should cover: (i) Project logic and design, (ii) M&E systems and capacity, (iii) Evaluation questions of concern to stakeholders, (iv) Possible evaluation designs.

4.2 What major issues should be examined?

4.5 This section focuses primarily on the contents of 10% of the documents in the bibliography that provide official guidance on Evaluability Assessment, which relates to eight international organisations.

4.6 The UNIFEM "Guidance Note on Carrying Out an Evaluability Assessment" (Sniukaite, 2009) uses three main categories, which have since been adopted by other organisations in their guidance (CDA, EC-Evalsed, and UNODC). They are:

- **The adequacy of the program design**, including its clarity, coherence, feasibility and relevance. This addresses "in-principle" evaluability, mentioned earlier.
- **The availability of information**, including both contents available and systems for making it available. This addresses "in-practice" evaluability, mentioned earlier.
- **The conduciveness of the context**, including stakeholders views and resources available. This addresses both "in-practice" evaluability and the utility of an evaluation.

4.7 The USAID guidance (Dunn, 2008) asks three broad questions which are similar in focus to the issue headings used by UNIFEM:

- **Is it plausible to expect impacts?** Do stakeholders share a clear understanding of how the program operates and are there logical links from program activities to intended impacts?
- **Is it feasible to measure impacts?** Is it possible to measure the intended impacts, given the resources available for the impact assessment and the program implementation strategy?
- **Would an impact assessment be useful?** Are there specific needs that the impact assessment will satisfy and can it be designed to meet those needs?

4.8 The ILO guidance is more narrowly focused on meeting the needs of their Results Based Management approach, by focusing on measurability. The topic headings being: Objectives, Indicators, Baseline, Milestones, Risk Management and M&E system.

UNFPA ToR for Evaluability Assessments are oriented by their Results Based Management approach, with the main headings asking about "Logical Sequence of the Chain of Results", "Indicators", "Means of Verification" and "Risks and Assumptions".

4.9   The IADB guidance (Soares, et al, 2010) has a similar focus on measurement, one which is expressed in their own particular definition of evaluability, as "*the ability of an intervention to demonstrate in measurable terms the results it intends to deliver*". Evaluability inquiries focus on two sets of "dimensions", which seem to correspond to the first two of the three issues identified by UNIFEM:

- Substantive dimensions, which include problem diagnosis, project objectives, project logic, and risks
- Formal dimensions, which include Outcome indicators, Output indicators, Baselines for outcomes, Baselines for outputs, and Monitoring and Evaluation systems and resources

4.10   The third issue of conduciveness of context for an evaluation is understandably absent since in the IADB Evaluability Assessment they are carried out at a very early stage, when a project is being considered for approval.

4.11   While there is no AusAID guidance on Evaluability Assessment the Evaluability Assessments that have been carried out in Indonesia have been oriented by the draft M&E standards developed for that country program. These contain six standards, one of which focuses on quality of project design and five which focus on monitoring and evaluation processes. This approach has similarities with the IADB process both in its focus and also in its closer connection to wider quality assurance processes. Individual consultants have had freedom to explore other dimensions of evaluability, including in at least one of the three cases reviewed, an assessment of constraints and opportunities for utilisation i.e. UNIFEM's third issue.

4.12   Recent Dutch guidance for the NDC pays more attention to implementation than design (Ruben 2012). The main questions being: 1. Does the program serve the population for whom it was designed? 2. Does the program have the resources (available/used) as scheduled in the program design? 3. Are the program activities being implemented as designed? 4. Does the program have the capacity to provide data for an evaluation?

4.13   In contrast to USAID, ILO, IADB and AusAID, the draft IDRC guidance (Monk, 2012b) pays much more attention to the evaluation context and design, some attention to program design and very little to the issue of information availability. The IDRC guidance is the least prescriptive.

4.14   The EBRD discussion paper on Evaluability Assessment also suggests examining context, by asking if attention has been paid to identifying important risks i.e. reasons

why the project design may not work as expected. This can be seen as another perspective on feasibility, usually considered under project design. Risk is also a topic covered by the ILO Evaluability Assessment tool.

> Recommendation 10: An examination of guidance documents produced by eight international agencies suggests that an Evaluability Assessment should attend to three broad types of issues:
> - The program design
> - The availability of information
> - The institutional context
>
> These relate closely to three purposes of Evaluability Assessment discussed earlier (page 8)
>
> The division of attention across these areas will be subject to the timing of an Evaluability Assessment, with design being the main focus at a quality assessment stage and information availability and conduciveness becoming relatively more important during implementation and immediately prior to an evaluation.

4.15   Within each of the three main issue areas there a number of specific criteria and associated questions that can be asked. These are summarised in the three tables that follow. Their content is a synthesis of questions used in a range of Evaluability Assessment tools reviewed during the review of the Evaluability Assessment.

1. Project Design (as described in a Theory of Change, Logical Framework or narrative)	
Clarity?	Are the long-term impact and outcomes clearly identified and are the proposed steps towards achieving these clearly defined?
Relevant?	Is the project objective clearly relevant to the needs of the target group, as identified by any form of situation analysis, baseline study, or other evidence and argument? Is the intended beneficiary group clearly identified?
Plausible?	Is there a continuous causal chain, connecting the intervening agency with the final impact of concern?
	Is it likely that the project objective could be achieved, given the planned interventions, within the project lifespan? Is there evidence from elsewhere that it could be achieved?
Validity and reliability?	Are there *valid* indicators for each expected event (output, outcome and impact levels)? i.e. will they capture what is expected to happen? Are they *reliable* indicators? i.e. will observations by different observers find the same thing?
Testable?	Is it possible to identify which linkages in the causal chain will be most critical to the success of the project, and thus should be the focus of evaluation questions?
Contextualised?	Have assumptions about the roles of other actors outside the project been made explicit (both enablers and constrainers)? Are there plausible plans to monitor these in any practicable way?
Consistent?	Is there consistency in the way the Theory of Change is described across various project multiple documents (Design, M&E plans, work plans, progress reports, etc.)?
Complexity?	Are there expected to be multiple interactions between different project components [complicating attribution of causes and identification of effects]? How clearly defined are the expected interactions?
Agreement?	To what extent are different stakeholders holding different views about the project objectives and how they will be achieved? How visible are the views of stakeholders who might be expected to have different views?

4.16   Commentary: The above list leaves out many criteria that readers may think is indicative of a "good" ToC, e.g. alignment with current policy objectives or a gender analysis informed strategy. However a "good" ToC and an evaluable ToC is not necessarily the same thing. A ToC may be evaluable because the theory is clear and plausible, and relevant data is available. But as the program is implemented, or following its evaluation, it might be discovered that the ToC was wrong, that people

or institutions don't work the way the theory expected them to do so i.e it was actually a "bad" ToC. Alternately it is also possible that a ToC may turn out to be "good", but the poor way it was initially expressed made it un-evaluable, until remedial changes were made.

4.17   Ideally an Evaluability Assessment of the project design should take place before it is approved, as part of a wider Quality Assurance process. In reality the practical details of many project designs are articulated during inception periods and during implementation thereafter. In practice an assessment of the project design should be part of the Evaluability Assessment at any stage, regardless of whether evaluability was examined during project approval.

| 2. Information availability	
Is a complete set of documents available?	…relative to what could have been expected? E.g. Project proposal, Progress Reports, Evaluations / impact assessments, Commissioned studies
Do baseline measures exist?	If baseline data is not yet available, are there specific plans for when baseline data would be collected and how feasible are these? If baseline data exists in the form of survey data, is the raw data available, or just selected currently relevant items? Is the sampling process clear? Are the survey instruments available? If baseline data is in the form of national or subnational statistics, how disaggregated is the data? Are time series data available, for pre-project years?
Is there data on a control group?	Is it clear how the control group compares to the intervention group? Is the raw data available or just summary statistics? Are the members of the control group identifiable and potentially contactable? How frequently has data been collected on the status of the control group?
Is data being collected for all the indicators?	Is it with sufficient frequency? Is there significant missing data? Are the measures being used reliable i.e. Is measurement error likely to be a problem?
Is critical data available?	Are the intended and actual beneficiaries identifiable? Is there a record of who was involved in what project activities and when?
Is gender disaggregated data available?	In the baseline? For each of the indicators during project intervention? In the control group? In any mid-term or process review?
If reviews or evaluations have been carried out…	Are the reports available? Are the authors contactable? Is the raw data available? Is the sampling process clear? Are the survey instruments available?
Do existing M&E systems have the capacity to deliver?	Where data is not yet available, do existing staff and systems have the capacity to do so in the future? Are responsibilities, sources and periodicities defined and appropriate? Is the budget adequate?

3. Institutional context	
Practicalities	
Accessibility to and availability of stakeholders?	Are there physical security risks? Will weather be a constraint? Are staff and key stakeholders likely to be present, or absent on leave or secondment? Can reported availability be relied upon?
Resources available to do the evaluation?	Time available in total and in country? Timing within the schedule of all other activities? Funding available for the relevant team and duration? People with the necessary skills available at this point?
Is the timing right?	Is there an opportunity for an evaluation to have an influence? Has the project accumulated enough implementation experience to enable useful lessons to be extracted? If the evaluation was planned in advance, is the evaluation still relevant?
Coordination requirements?	How many other donors, government departments, or NGOs need to be or want to be involved? What forms of coordination are possible and/or required?
Demands	
Who wants an evaluation?	Have the primary users been clearly identified? Can they be involved in defining the evaluation? Will they participate in an evaluation process?
What do stakeholders want to know?	What evaluation questions are of interest to whom? Are these realistic, given the project design and likely data availability? Can they be prioritised? How do people want to see the results used? Is this realistic?
What sort of evaluation process do stakeholders want?	What designs do stakeholders express interest in? Could these work given the questions of interest and likely information availability, and resources available?
What ethical issues exist?	Are they known or knowable? Are they likely to be manageable? What constraints will they impose?
What are the risks?	Will stakeholders be able to manage negative findings? Have previous evaluation experiences prejudiced stakeholder's likely participation?

4.18   **Commentary**: Evaluation questions are of particular interest to some commissioners of Evaluability Assessments, such as DFID, where their examination is one of six objectives for recent desk-based Evaluability Assessments. Relevant evaluation questions can be identified by examining the project design, and explicating the built-in hypotheses about what will work (1. above). Evaluation questions of interest are also likely to emerge through consultations with stakeholders (3. above). Without pre-

empting the design of an evaluation, they can then inform the assessment of the availability of information (2. above).

4.19   Evaluation designs need to be explored by an Evaluability Assessment *in as much as* stakeholders have expressed interest in or preferences for specific approaches. Are these practically possible given the context and appropriateness of the project design, the questions being asked and the likely availability of data?. The Evaluability Assessment should however avoid beginning an evaluation design process. That should be the responsibility of other parties who are contracted to undertake the evaluation.

4.20   This diagram attempts to summarise the relationships between the aspects of an Evaluability Assessment described in the tables above, and how they relate to evaluation design:
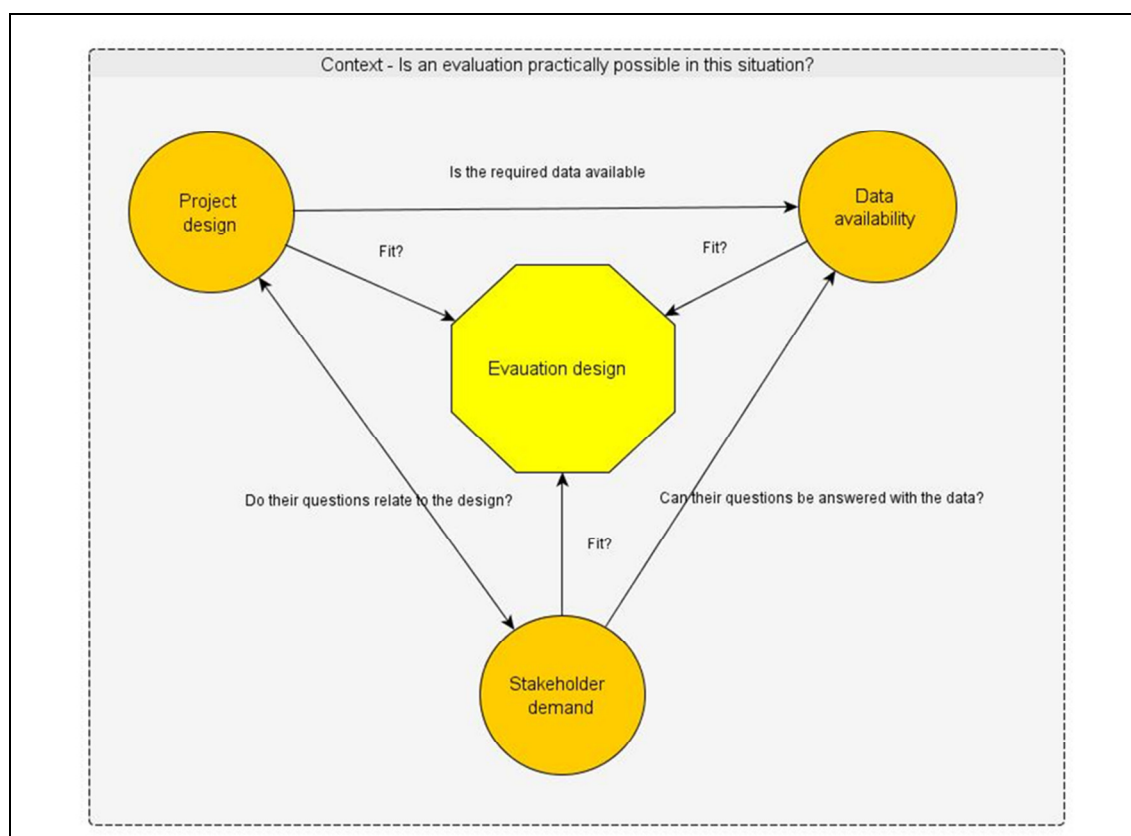
4.3   Why use checklists?

4.21   Checklists are a means of ensuring the systematic examination of all relevant issues, across all projects being examined. More than half (11) of the 19 agencies found to be using Evaluability Assessments have used checklists in one form or another. Checklists can be used as a standalone tool or incorporated as questions into the Terms of Reference for an Evaluability Assessment whose results are expected to be summarised in a substantial written report (e.g. UNFPA, 2012).

4.22   Checklists can vary in content and use. At one extreme is the draft IDRC checklist, which is essentially a suggested agenda for discussion with limited requirements on how the results of each discussion will be documented. In contrast, the M&E standards used in the AusAID Evaluability Assessments have some authority and their applicability may not be negotiable.

4.23   The UNIFEM checklist has 18 questions which seem to require a binary yes/no answer (Sniukaite 2009). This should be relatively simple to use, perhaps too much so. UNFPA has a similar checklist, but with space for explanatory comments (UNFPA 2012).

4.24   Rating scales are used by the ILO and IADB. Rating scales can have supporting guidance on their use. This enables consistency in use by different assessors. IADB has made use of such scales, with each point on a three point scale supported by an example (Office of Evaluation and Oversight 2000). The ILO Evaluability Assessment Tool also has supporting advice on the use of its rating scales.

4.25   Checklists can be structured into sections and sub-sections, enabling meso-level judgements to be built up from micro-judgements. Examples can be seen in the checklists used by ILO and IADB.

4.26   Checklists vary in length. The ILO Evaluability Assessment Tool has 23 separate questions, in six groups. UNIFEM has 18 questions in three groups. The IADB has 29 questions in eight groups. The UNODC has 28 questions in three groups.

4.27   The three tables presented above are a form of checklist, which could be further developed and customised.

> Recommendation 11: Evaluability Assessment checklists should be used. They encourage comprehensive coverage of relevant issues, and visibility of those that are not covered. They can be used as stand–alone tools along with ratings, or be supported by comment and analysis or have a more background role informing the coverage of a detailed narrative report.

4.4   Why calculate aggregate scores?

4.28   The aggregation of individual judgements into a total score can enable comparisons of evaluability across projects and across time. Without that capacity it will be more difficult to accumulate lessons about what is working, or not.

4.29   The IADB has used evaluability scores to make systematic comparisons of project designs across different evaluability dimensions and across different project types, on three occasions since 2001. In 2009 the IADB noticed an overall decline in evaluability thought to be associated with a substantial increase in the number of projects being funded. This may have influenced their decision to switch from three yearly

examination of evaluability to an annual process examining a random sample of projects.

4.30   The 2001 study of human rights projects by Poate et al included an extensive analysis of the frequency of different kinds of evaluability issues in the 28 projects they examined. Further analyses of such data sets can also help identify what aspects of evaluability are the best predictors of overall evaluability status. For example, it appears that a combination of two of the 13 criteria used can be used to identify 86% of the most evaluable projects[9]. Further analysis could shed light on the importance of contextual factors (e.g. each project's country and sector) in determining evaluability.

4.31   The ILO Evaluability Assessment Tool does aggregate scores across all six dimensions of evaluability. It is not known whether there has been any analysis of these scores.

4.32   Aggregate scores can be combined with minimal score requirements to deliver judgements about what needs to be done. The IADB requires that a "minimum evaluability threshold of 5 will be required for all operations to be submitted to the Board of Executive Directors". The UNODC Evaluability Assessment template requires scores on four separate checklists to exceed an average 50% before an evaluation can take place (Gunnarsson 2012). How this works in practice is not yet known because the template is not yet in use by UNODC project managers.

4.33   The ILO Evaluability Assessment Tool has four grades of evaluability, based on the aggregate scores. The lowest grade is "Not evaluable". It is not yet clear if since the tool has been put into operation whether any such judgements have been made, and what the consequences have been.

"Wholey's Evaluability Assessment (EA) framework provided very helpful guidelines for the current assessment. However, attempting to adapt his framework was no easy task. A recurrent issue concerns the idea of thresholds for what qualifies enough versus too little evidence to consider a condition met. For example, how much evidence do we need before assuming that the goals of the evaluation strategy are appropriate and feasible? … Seeking through the literature to see how other researchers have dealt with this issue proved to be futile because to our knowledge no authors have addressed this point. (D'Ostie-Racine, Dagenais, and Ridde 2013)

Recommendation 12: The aggregation of individual judgements into a total score is good practice because it enables comparisons of evaluability across projects and across time, and thus lessons learned from differences in evaluability.

The use of minimum threshold scores is not advisable unless there are very goods grounds for defining such a threshold.

---

[9] (a) Identifiable outputs" , (b) "How easily can benefits be attributed to the project intervention alone". See
http://mandenews.blogspot.co.uk/2013/04/an-example-application-of-decision-tree.html

4.5 Why use weightings?

4.34 Any aggregation of scores on a checklist involves assumptions about their relative importance of different items. If there is no explicit weighting of items then all items are in effect being treated as being equally important, which is unlikely to be the case in practice.

4.35 Most of the checklists identified during this review do not seem to have any explicit weighting system. One exception is the ILO Evaluability Assessment Tool, where a different weighting is given to the scores on each of six dimensions (ILO Evaluation Unit 2012). According to the ILO guidance "The weight/ratio defined by the tool is based on the expertise, experiences, and best practices of EVAL". An aggregate score is then generated by multiplying the dimension score by its weighting. Within each dimension there are multiple questions, the qualitative answers to which are used to derive, by expert judgement, a score for the dimension as a whole.

4.36 It should be noted that weighting values do not need to be fixed into the initial design of a checklist. They can be assigned along with the performance judgements, as part of each Evaluability Assessment. Doing so would enable the Evaluability Assessment to be more context sensitive. No examples were found of this approach.

> Recommendation 13: Where scored checklists are used there should be explicit weightings, to avoid mistaken assumptions about all things being equally important. Weightings can be either built in by the checklist designer or provided by the checklist user – as part of their assessment. If possible, explanations should be sought for weightings, in order to make judgements more transparent.

4.6 Can evaluability really be measured?

4.37 The development of an instrument to assess evaluability in the OECD-DAC sense seems to imply belief in the existence of a preferred or desirable form of evaluation. This seems questionable given the numerous schools of evaluation that seem to exist. It would be more reasonable if the agency undertaking Evaluability Assessments has a specific view on what forms of evaluation are desirable. For example, some agencies like the ILO are quite explicit in their orientation. Their Evaluability Assessment guidance begins by noting that "The ILO is committed to strengthening the Office-wide application of results-based management". Similarly, the IADB defines evaluability as "ability of an intervention to demonstrate in measurable terms the results it intends to deliver" (Soares et al. 2010).

4.38 Others like IDRC are explicit in their view that an Evaluability Assessment should not constrain the type of evaluation that can be carried out. With this in mind Monk has taken a more radical position and argued that clarity on program theory, the existence of SMART indicators and the presence of baseline data are not essential prerequisites for all types of evaluation. She points out, for example, that a program theory is not

needed for a Goal Free evaluation. Most other agencies are less explicit about the range of evaluation methods their Evaluability Assessment guidance relates to.

4.39   If the use of a range of evaluation methods is acceptable then this raises questions about the usefulness of any fixed checklist, and even more so, the use of a scoring system based on the results of the checklist, especially one with pre-defined cut-off points below which a project is seen as un-evaluable. For example, those used by UNODC, ILO, IADB, and ITAD for SIDA.

4.40   If checklists and scores are in doubt then what are the alternatives? Monk (2012) has proposed that a worksheet of questions would "*be used by a member of the [IDRC] Evaluation Unit in their discussions with the Program Officer who is commissioning the evaluation. The Evaluability Assessment would be presented as a conversation… The questions in the worksheet are meant to guide the conversation. The representative from the Evaluation Unit will rephrase, add and drop questions as they see fit.*"

4.41   The risk here is of a process that is applied with variable degrees of thoroughness across a range of different projects, with very little comparability of results. While it may aid improvements in evaluability on a case by case basis, it would not enable any prioritisation in the commissioning of evaluations or systematic learning about where the most common problems were to be found.

4.42   There is an alternative. Scored checklists could still be useful if they were seen as a systematic way of generating an explicit assessment of *likely challenges* facing an evaluation. Not a final judgement on evaluability. The onus would then be on other parties to explain how they would respond to these challenges. These other parties could be project managers who are expected to develop functioning M&E frameworks, or consultants bidding for an evaluation, who would need to explain in their proposals how they would address the identified challenges through their choice of evaluation approach.

4.43   The use of aggregate scores would be a useful *index of difficulty* signalling the relative scale of the challenges faced. So would the use of weightings being assigned to different categories of challenges, during each Evaluability Assessment[10].

> Recommendation 14: The results generated by a scored checklist should be seen as *an index of difficulty*, which then needs to be responded to by program managers and/or evaluators when they are commissioning or planning an evaluation. Not as a final judgement on evaluability, given the range of evaluation methods and purposes that exists.

---

[10] As distinct from building in a standardised weighting to be applied across the board to all projects subject to an Evaluability Assessment.

4.7 What outputs can be expected?

4.44   These proposals have been informed by a review of Evaluability Assessments by DFID, USAID, UNICEF, UNIFEM, UNFPA, and AusAID.

4.45   Two types of Evaluability Assessment outputs might be expected. The first are *assessments*, which relate to the purposes of the Evaluability Assessment spelled out earlier in this paper, concerning:
- The *evaluability* of a project, given its design and the information that will be available
- The *practicality and utility* of an evaluation, given the nature of the project and the context in which an evaluation could take place

The second is *recommendations* concerning:
- Changes to the project design, which will make it more evaluable
- Development of the associated M&E systems, which will make it more evaluable
- Aspects of an evaluation design, which would help ensure its usefulness
  - Timing – if and when an evaluation would usefully take place
  - Evaluation questions relevant to potential users of an evaluation and to the design of the project
  - Evaluation methods and designs relevant to the evaluation questions and the availability of information

4.46   Some Evaluability Assessment ToRs go further and also request proposals concerning appropriate budgets and relevant expertise for an evaluation (DFID, USAID, JSI), and even timeframes and milestones. Recent DFID ToR for Evaluability Assessments place more emphasis on assessing evaluation questions and designs than other agencies, reflecting the importance they are now given in the Business Cases, compared to earlier forms of project proposals that were used.

4.47   However, expectations should be bounded. Recommendations should inform the design of ToR for an evaluation, but not pre-empt the design of an evaluation, especially if the Evaluability Assessment has been carried out quite early in a project cycle.

4.48   An outline structure for ToRs for an Evaluability Assessment has been provided in Annex G.

Recommendation 15: Outputs of an Evaluability Assessment should include both assessments and recommendations. Assessments should cover (a) evaluability of the project, referring both to the project design and the availability of information, and (b) the practicality and utility of an evaluation. Recommendations can refer to: (a) changes in project design and associated M&E systems to make it more evaluable, (b) options for evaluation timing, evaluation questions and evaluation methods, to help ensure the usefulness of an evaluation.

Recommendations should inform the design of ToR for an evaluation, but not pre-empt the design of an evaluation.

5   CONCERNS

5.1   Does it work?

5.1   The IADB is the only organisation found in this research to have extended experience in Evaluability Assessment and the only one known to have assessed that experience. The IADB Office of Evaluation and Oversight has reviewed three rounds of Evaluability Assessments, of all of its projects approved in the years 2001, 2005 and 2009. The impact of the Evaluability Assessments on overall project design quality seems to have been limited at best (Office of Evaluation and Oversight, 2010). The 2006 project designs showed some improvements on those in 2001 but subsequently the 2009 project designs were rated worse than those in 2001 on eight out of nine of the dimensions of evaluability. However, these changes have coincided with a doubling in the amount and number of project approved. Thus it could be argued (in the absence of any control group) that without the Evaluability Assessments the design quality in 2009 may have been worse still. The IADB has since moved over to a yearly process, involving a random sample of one third of all projects. It has also revised its quality assurance processes associated with project design and approval (Office of Evaluation and Oversight 2011). The Evaluability Assessment now functions as a validity check of self-assessment processes.

5.2   There are other forms of data on the effectiveness of Evaluability Assessments, but this has not been systematically collated and assessed in relation to international development agencies. Probably the most noticeable form of impact is where decisions are made not to proceed with a proposed evaluation. In the Evaluability Assessment of a USAID funded micro-enterprise program in Brazil it was decided not to proceed with plans for an impact evaluation (Snodgrass, Magill, and Chartock 2006). In other cases the impact may be in the form of delays or reduced level of ambition or scope, rather than cancellations, such as the evaluation of DFID's Empowerment and Accountability and Gender vision portfolio of projects (Davies et al. 2012).

5.3   Reviews of the experience of using Evaluability Assessments in the United States in the 1970's and 1980s suggest that subsequent delays, if not deferrals, were quite common. In reviewing the implementation of the method, Rog found that "Most of the studies provided options to improve program management, develop performance measures, and design evaluation strategies. Few were followed by subsequent evaluations, however" (Leviton et al. 2010).

5.4   In the light of what may seem to be limited evidence about the effectiveness of Evaluability Assessments what justification is there for doing one? The answer depends on two factors: (a) the cost of an Evaluability Assessment relative to the cost of an evaluation, (b) the extent to which an Evaluability Assessment subsequently improves an Evaluation. The smaller the proportionate cost, the smaller the increment in evaluation quality that is needed to justify the cost. If an Evaluability Assessment is 10%

of the cost of an evaluation then a modest 10% improvement in the value of the subsequent evaluation would already cover the cost of the Evaluability Assessment.

> <u>Recommendation 16</u>: While there is limited systematic evidence on the effectiveness of Evaluability Assessments the relatively low costs of Evaluability Assessments means that they only need to make modest improvements to an evaluation before their costs can be recovered.

5.2 What are the risks?

5.5 There are number of risks associated with Evaluability Assessments, all of which can be found with many other organisational procedures. They include:

- **Conflation and confusion** of purpose. One problem already seen in some Evaluability Assessment instruments is conflation of the assessment of the extent to which a project's strategy aligns with an organisation's wider policy objectives, with the assessment of its feasibility and measurability. An IADB review has also noted the need to separate out risk management scoring from their Evaluability Assessment (Office of Evaluation and Oversight 2011). Confusion can also exist amongst those contacted by an Evaluability Assessment team. In their Evaluability Assessment of the Sida project, Poate et al (2000) noted "the concept of evaluability was an unfamiliar one, and many interviewees still believed the exercise was in fact a kind of Sida evaluation, and they treated it as such".

- **Evaluation overload:** This is more likely where the Evaluability Assessment segues into an evaluation as a result of extensive consultations with project stakeholders. Desk based assessments are much more limited in scope and in their demands on stakeholders.

- **Delay**: Imas and Rist (2009) warn that it can unnecessarily delay evaluations if it is applied to all planned evaluations. IADB has managed this risk by using a sampling strategy. Where Evaluability Assessments are voluntary, program managers can make their own assessment of acceptable delays.

- **Additional cost burden**: The information that is available on Evaluability Assessment costs suggests that the net addition to cost, on top of evaluation costs is not likely to be substantial. It is possible that the cost of Evaluability Assessments may not be recovered through improved evaluation quality.

- **Ineffectiveness**: Institutionalising Evaluability Assessment may lead to it becoming a formality with no consequences. An IADB document reports "As a note of caution, the report presents some evidence that even SG projects with high DEM scores may not ultimately be evaluable if project teams do not have adequate incentive to follow up on monitoring and evaluation needs post-approval. Specifically, OVE's review of post-approval Loan Contracts and Loan Results Reports for the projects reviewed do not indicate that evaluability aspects missing at approval were later addressed as intended". (Office of Evaluation and Oversight 2011)

5.6   Cautionary notes are not found in many guidance documents on Evaluability Assessment. However, Evalsed (2009) does provide the following useful advice:

"The strengths of this approach is that it has the potential to improve programmes and their performance and ensure that only those evaluations that are likely to justify the efforts involved, actually take place. For these strengths to be realised the approach:

- Needs to be applied with a light touch. The approach should be seen as a quick, low cost and time limited intervention built into the management functions.
- Expectations should be realistic and not too ambitious. A misapprehension to be avoided is that such an exercise can deliver certainty or that Evaluability Assessment can substitute for other evaluations also likely to be needed.
- Should be applied selectively. Not every programme would benefit from such an approach and programme managers would need to develop their own criteria for when it is worthwhile, e.g. when the partners in a programme are open to change, when there is a prior uncertainty about the form of evaluation to implement etc."

5.7   One risk not seen in the literature on Evaluability Assessment is the political risk of an Evaluability Assessment unpacking and challenging a project design, especially one that has been approved or is close to approval. One of the reasons why Evaluability Assessments have not been used may be the fear of such challenges.

Recommendation 17: The biggest risk of failure facing an Evaluability Assessment is likely to be excessive breadth of ambition: reaching into evaluation design or evaluation itself. This risk may be higher when Evaluability Assessments are undertaken in-country in association with stakeholders, versus at a distance via a desk based analysis. It should also be recognised that Evaluability Assessment may be seen as challenging, if there are already doubts about a project design.

5.3   Why has Evaluability Assessment not been part of standard practice?

5.8   Themessl-Huber (2010) has listed eight possible reasons, four of which seem relevant in the development aid context:
- Evaluability Assessments are under-reported in the published literature. The production of the bibliography associated with this synthesis report may help address this problem, by widening knowledge of what has been done to date.
- Evaluability Assessments are seen as an extra expense, which takes additional time. The analysis of time and costs involved, given above, suggests that this should not be a major concern, especially when compared to the average duration and cost of evaluations of medium and large scale projects.
- Lack of a clearly defined methodology. This is becoming less so. In the last three years there has been a burst of publications by international organisations aimed at spelling out how Evaluability Assessments should be carried out. Recommendations made in this synthesis paper should also help.

- Evaluability Assessments which are undertaken are seen as a preliminary step in evaluation process rather than as an independent tool. The experience referred to in this synthesis paper suggests that earlier use of Evaluability Assessment is more valuable, in terms of potential to improve project design as well as inform the design of their evaluation.

Recommendation 18: There has been a resurgence in the use of Evaluability Assessment but not yet in the published literature on Evaluability Assessment. Guidance material is becoming more available but reviews of the use of Evaluability Assessments are still scarce The online bibliography produced as a part of this synthesis report should be periodically updated and publicised to make sure experiences with Evaluability Assessment are widely accessible, and open to further reviews.

Annex A: References cited in this report

Please note that the complete bibliography of 131 references can be found online here:
http://mande.co.uk/blog/wp-content/uploads/2013/02/Zotero-report.htm

D'Ostie-Racine, Léna, Christian Dagenais, and Valéry Ridde. 2013. "An Evaluability Assessment of a West Africa Based Non-Governmental Organization's (NGO) Progressive Evaluation Strategy." *Evaluation and Program Planning* 36 (1) (February): 71–79. doi:10.1016/j.evalprogplan.2012.07.002.

Davies, Rick, Sarah-Jane Marriot, Gibson, and Emma Haegeman. 2012. "Evaluability Assessment For DFID's Empowerment and Accountability And Gender Teams". IDL.

Dawkins, Nicola. 2010. "Nicola Dawkins on Evaluability Assessment and Systematic Screening Assessment · AEA365." *AEA365 | A Tip-a-Day by and for Evaluators.* http://aea365.org/blog/?p=1005.

Dunn, E. 2008. "Planning for Cost Effective Evaluation with Evaluability Assessment". USAID. http://pdf.usaid.gov/pdf_docs/PNADN200.pdf.

Epstein, Diana, and Jacob Alex Klerman. 2012. "When Is a Program Ready for Rigorous Impact Evaluation? The Role of a Falsifiable Logic Model." Evaluation Review 36 (5) (October 1): 375–401. doi:10.1177/0193841X12474275.

Evalsed, European Commission. 2009. "Evaluability Assessment." *Source Book: Methods and Techniques.* http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/sourcebooks/method_techniques/structuring_evaluations/evaluability/index_en.htm.

Gunnarsson, Charlotte. 2012. "Evaluability Assessment Template". UNODC. http://www.unodc.org/documents/evaluation/IEUwebsite/Evaluability_Assessment_Template.pdf.

Imas, Linda G. Morra, and Ray C. Rist. *The Road to Results: Designing and Conducting Effective Development Evaluations.* World Bank Publications, 2009.

ILO Evaluation Unit. 2012. "Dimensions of the Evaluability Instrument". ILO. http://www.ilo.org/wcmsp5/groups/public/---ed_mas/---eval/documents/publication/wcms_165985.pdf.

IOD PARC. 2011. "UNIFEM Strategic Plan 2008-2011 Evaluability Assessment". UNIFEM. http://www.unwomen.org/wp-content/uploads/2012/03/EvaluationReport-UNIFEMEvaluabilityAssessment_en.pdf.

Jung, S. M. 1980. "Implementation of the Career Education Incentive Act. First Interim Report on the Evaluability Assessment." http://www.eric.ed.gov/ERICWebPortal/recordDetail?accno=ED186679.

Leonard, Keith, and Amelie Eulenberg. 2012. "Evaluability – Is It Relevant to EBRD?" European Bank for Reconstruction and Development. http://www.ebrd.com/downloads/about/evaluation/130305Evaluability.pdf.

Leviton, Laura C., Laura Kettel Khan, Debra Rog, Nicola Dawkins, and David Cotton. 2010. "Evaluability Assessment to Improve Public Health Policies, Programs, and Practices

★.” *Annual Review of Public Health* 31 (1): 213–233.
doi:10.1146/annurev.publhealth.012809.103625.

LMIT–BPS. 2008. “Evaluability Assessment of Pawnbroker Databases”. U.S. Department of
Justice. https://www.ncjrs.gov/pdffiles1/nij/pawnbroker-databases.pdf.

Monk, Heidi. 2012. “Evaluability Assessment in Theory and Practice.” IDRC.

Monk, Heidi. 2012b. “Evaluability Assessment at IDRC”. IDRC.

Office of Evaluation and Oversight. 2000. “Working Paper: Evaluability Assessment in Project
Preparation”. IADB.
http://idbdocs.iadb.org/wsdocs/getdocument.aspx?docnum=36531372.

———. 2010. “Evaluability Review of Bank Projects 2009”. IADB.
http://idbdocs.iadb.org/wsdocs/getdocument.aspx?docnum=35594535.

———. 2011. “Approach Paper: Evaluability Review of Bank Projects 2011”. IADB.
http://idbdocs.iadb.org/wsdocs/getdocument.aspx?docnum=37107743.

Ogilvie, David, Steven Cummins, Mark Petticrew, Martin White, Andy Jones, and Kathryn
Wheeler. 2011. “Assessing the Evaluability of Complex Public Health Interventions:
Five Questions for Researchers, Funders, and Policymakers.” *Milbank Quarterly* 89 (2):
206–225. doi:10.1111/j.1468-0009.2011.00626.x.

Plowman, Beth, and Henry Lucas. 2011. “Evaluability Study of Partnership Initiatives
Norwegian Support to Achieve Millennium Development Goals 4 & 5 -”. NORAD.
http://www.norad.no/en/tools-and-
publications/publications/publication/_attachment/237803?_download=true&_ts=12e
48e4a647.

Poate, D., R. Riddell, Tony Curran, and Nick Chapman. 2000. “The Evaluability of
Democracy and Human Rights Projects Volume 1”. SIDA.
http://www.odi.org.uk/sites/odi.org.uk/files/odi-assets/publications-opinion-
files/2323.pdf.

Ruben, Ruerd. 2012. “Evaluablity Assessment - Preparatory Steps Before Starting an
Evaluation”. Policy and Operations Evaluation Dept. , Development Cooperation,
Ministry of Foreign Affairs of the Netherlands.
http://www.oecd.org/development/evaluationofdevelopmentprogrammes/dcdndep/4
6436210.ppt.

Sniukaite, Inga. 2009. “Guidance Note Carrying out an Evaluability Assessment”. UNIFEM.
http://erc.undp.org/unwomen/resources/guidance/Guidance%20Note%20-
%20Carrying%20out%20an%20Evaluability%20Assessment.pdf.

Snodgrass, Don, John Magill, and Andrea Chartock. 2006. “Evaluability Assessment of the
USAID/Brazil Micro and Small Enterprise Trade-Led   Growth Program”. USAID.
http://www.value-
chains.org/dyn/bds/docs/559/USAID%20Brazil%20Evaluability%20Assessment%2020
06.pdf.

Soares, Yuri, Alejandro Pardo, Veronica Gonzalez, and Sixto Aquino. 2010. “Ten Years of
Evaluability at the IDB”. IADB.
http://www.oecd.org/development/evaluationofdevelopmentprogrammes/dcdndep/4
6436272.ppt.

Themessl–Huber, Markus. 2010. "Evaluability Assessment: An Overview."
        http://www.docstoc.com/docs/34214980/Evaluability-assessment.

UNDP. 2007. "UNDP Job: 'Evaluability Assessment (EA) of UN Women Pacific Sub
        Regional Office (SRO) Annual Work Plan (AWP) and Programme Plans' Posted on
        the UN Job List." http://unjoblist.org/vacancy/?270757.

UNEG. 2008. "Evaluability Assessments of the Programme Country Pilots Delivering as One
        UN Synthesis Report". UNEG.

UNIFEM. 2012. "UNIFEM Jobs – 32934– Evaluability Assessment (EA) of UN Women
        Pacific S." *UNDP*. http://jobs.undp.org/cj_view_job.cfm?cur_job_id=32934.

UNFPA. 2012. "Terms of Reference of the Consulting Team for the UNFPA 7th Country
        Programme Baseline Data Validation Study". UNFPA.

Yantio, Debazou Y. 2008. "Evaluability Assessment of the Government of Tanzania and
        UNICEF Interventions in the Seven Learning Districts". UNICEF.

Annex B: The search process

The Evaluability Assessment bibliography is a result of the following:
- Searches via Google Scholar and Google Search to find documents with "evaluability" in the title. The first 100 items in the search result listing were examined.
- Searches via PubMed, JSTOR and Sciverse using the same keyword, and with the same limit within each search result.
- An inquiry made via the MandE NEWS, Xceval and Theory Based Evaluation email lists.
- Scanning of references within academic documents on evaluability found with high citation counts and within Evaluability Assessments and guidelines produced by international agencies
- References referred to by interviewees within international agencies.

The bibliography is limited to documents available prior to December 2012. It is available online at:

http://mande.co.uk/blog/wp-content/uploads/2013/02/Zotero-report.htm

The following where not included in the bibliography:

- Discussions of Evaluability Assessment within documents and books on evaluation e.g.
  - "DANIDA EVALUATION GUIDELINES." 2012. Danida. http://amg.um.dk/en/~/media/amg/Documents/Technical%20Guidelines/Evaluation/EVAL-guidelines-WEB.ashx.
  - The Handbook of Environmental Policy Evaluation. 2012. Routledge.
  - Vedung, Evert. 2000. Public Policy & Program Evaluation. Transaction Publishers.
  - Wholey, Joseph S., Harry P. Hatry, and Kathryn E. Newcomer. 2010. Handbook of Practical Program Evaluation. John Wiley & Sons.
- Evaluability Assessments carried out as the first stage of an evaluation and then included as an initial chapter in the report on the evaluation

Follow up interviews were held via skype with staff and/or consultants associated with these organisations:
- AusAID – 3
- DFID – 3
- GAVI – 1
- IADB – 1
- IDRC – 2
- NDC – 1
- USAID – 2
- UNEG – 1

Results of the search process were compiled using Zotero. [Zotero](#) is free and open-source reference management software to manage bibliographic data and related research materials (such as PDF files). References were coded by type within a separate Excel file.

Caveats concerning bar chart on page 6

Bear in mind this chart may also reflect the greater accessibility of the most recent documents, relative to older documents. Leviton et al (2010) reports that "In the late 1970s and 1980s, more than 50 Evaluability Assessments were conducted, 33 of these between 1980 and 1985. Wholey left government and use of the technique dropped off significantly. Between 1986 and 2000, only eight Evaluability Assessments could be identified". The search results may also under report the amount of Evaluability Assessment work being undertaken within the USA in the last decade. Leviton (2010) has identified 50 in the field of public health.

Annex C: The Terms of Reference

Terms of Reference : Synthesis of literature on Evaluability Assessments
Purpose

To produce a short practical note that summarises the literature on Evaluability Assessments, and highlights the main issues for consideration in commissioning an Evaluability Assessment.

Introduction

DFID is seeking a contractor to provide an analytical report that summarises and analyses the literature on Evaluability Assessments (EAs), identifying and synthesising lessons about what works and what is less useful in EAs, challenges in carrying out an EA, and issues to consider in commissioning and managing Evaluability Assessments.

The main purpose of the study is to set out the existing literature on Evaluability Assessments, drawing on theoretical papers and practical documentation, to assist DFID and other donors and commissioners of evaluations to improve their use and management of Evaluability Assessments. This will in turn improve the relevance, cost-effectiveness and quality of development evaluations commissioned.

Audience

The primary audiences for the report are global evaluation advisers and development practitioners involved in commissioning and carrying out evaluations and Evaluability Assessments.

Objectives and scope

The contractor is expected to deliver a short report (maximum of 15 pages, excluding annexes) that should identify and synthesise the existing literature on Evaluability Assessments. The work should inform DFID and other development agencies about what an EA can achieve, the scope of work, and issues to consider in commissioning EAs. The report should identify the main rationale for carrying out and using EAs, and identify current bottlenecks in their uptake and use.

The work has been broken down into two components.

i) Component One : Reviewing the literature and assessing approaches

The scope of work will include theoretical literature on the scope and purpose of EAs, and an assessment of the type and quality of a number of Evaluability Assessments. The EAs reviewed should represent a range of thematic areas, types of programmes and evaluation types.

The study should include the following:

a) Identifying and assessing different definitions of Evaluability Assessment

b) Developing clear inclusion criteria for assessing EAs to be considered within the report, to provide a range of experience from different agencies and types of Evaluation Assessment.

c) Reviewing the bibliography attached at Annex A; and additional publically available material where this will add significant value. For cost purposes, we would expect the bibliography to be the primary data source, with a light touch review of additional material to address contextual differences outlined in point b) above.

d) Dialogue with up to 4 development agencies using EAs.

e) Clear and analytical identification and synthesis of types of findings from EAs, and lessons learned about scope and process.

f) A systematic assessment and analysis of methods, techniques and approaches (including TOR and EA questions) for carrying out Evaluability Assessments.

g) An assessment of strengths and weaknesses of the different approaches to EAs. This should take into account the timing of the EA, the scope and cost of the EA, the type of programme, programme context, and type of evaluation / scope of evaluation being assessed.

h) Identification of issues for consideration in commissioning Evaluability Assessments, including scope, content, cost, process, timing, and use.

i) Recommendations for commissioning future Evaluability Assessments.

j) Review of existing guidance in the DFID Evaluation Guidance and recommendations for revised text.

k) Identification of development partners and evaluation practitioners interested in developing understanding of EAs.

l) Facilitation of a workshop on the findings and recommendations for commissioning EAs. Participants will include: other donors, NGOs, consultants and members of research consortia who play a role in the design, implementation, monitoring and evaluation of DFID programmes. The workshop will be hosted by DFID.

A decision on whether Component 2 works goes ahead will be taken after 31 March 2013.

ii) Component 2: External Dissemination

➢ Publication of an article on Evaluability Assessments in a reputable and practice oriented journal.
➢ Presentation materials relating to the article for use at evaluation events.
➢ Presentation on the report and participation in any pre-agreed dissemination/communication events

Deliverables

The following deliverables are expected as part of the project:

Before 31 March 2013:

▪ An inception report/analytical framework for the report (of no more than 4 pages) to include an analytical framework for the Report and proposed workplan. This is expected to include an inclusion criteria of what will be covered as well as draft timeline for activities. **Due by 30 November 2012**.
▪ Draft Final report. **Due by 15 March 2013**. DFID will send comments on the draft report by **29 March 2013**.

- Facilitation and presentation at a workshop for external partners.
- Final report, taking on board suggestions to the draft final report. Due by **12 April 2013**. The final report should not exceed 15 pages, excluding annexes.
- Recommended changes to DFID Evaluation Guidance on EAs. Due by **12 April 2013**.

Before September 2013:

- An article on EAs submitted to a reputable and practice oriented journal.
- Presentation materials for use at evaluation events.
- Presentation on the report and participation in any pre-agreed dissemination/communication events.

Detail on the three final components will be discussed and agreed with DFID after submission of the final report.

Methods

The analysis and conclusions contained in the report should be based on the following:

- Desk review of a selected number of Evaluability Assessments from a range of international development agencies
- Desk review of available literature on Evaluability Assessments, drawing on the bibliography at Annex A
- Interviews and fact checking with relevant staff from DFID, and bilateral and multi-lateral agencies (e.g. AusAid, Danida, UNICEF, World Bank)

Contracting Arrangements and Timeframe[11]

This contract will be milestone based, with payment based on delivery of key outputs which must be completed and agreed by **12 April 2013**.

- o Inception report   10%
- o Draft report and facilitation of workshop   60%
- o Final report and review of DFID Guidance   30%

The study manager for technical issues will be Lina Payne (l-payne@dfid.gov.uk), all contracting issues (including amendments to deliverable dates and schedule of prices) will be dealt with by John Murray (j-murray@dfid.gov.uk). The successful consultancy is expected to undertake an internal QA product process prior to submission to DFID.

Additional scope and budget will be agreed after 12 April 2013, for attendance and presentation at international evaluation event(s); and peer reviewed article and presentation materials.

Final

1 November 2012

---

[11] This TOR builds on earlier work on EAs carried out as part of an evaluability assessment commissioned by DFID. The timeframe assumes a good level of existing knowledge of the literature.

Annex D: Examples of stage models in Evaluability Assessments

CDA (Reimann, 2012)	• Define focus, purpose, boundaries of and responsible staff and stakeholders involved in an EA. • Identify, review and analyze program documentation. • Identify and interview main stakeholders, including those responsible for program implementation and assumed beneficiaries. • Clarify program logic/theory of change/results chain. • Determine plausibility of program. • Draw conclusions and make recommendations if a program is ready for formal evaluation, what needs to be changed and/or what might be alternative evaluation designs.
UNODC (2012)	• Review of programme documentation. • Analysis of the information system defined in the programme (or related to the programme) and determination the information needs. • Interview of the main stakeholders. • Analysis of the programme.
NDC (Ruben 2012, after Smith, 1989)	• Identify relevant stakeholders. • Define boundaries of the program. • Analyze available program documents. • Clarify intervention theory (goals, resources, activities, outcomes). • Analyze stakeholders perceptions of the program. • Assess target population(s). • Discuss differences in outcome perceptions. • Determine plausibility of intervention model. • Discuss validity of the program. • Decide about continuation (= full evaluation).
UNIFEM (2009)	• Involving the intended users of evaluation information. • Clarifying the intended program. • Exploring program reality. • Reaching agreement on needed changes in activities or goals. • Exploring alternative evaluation designs. • Agreeing on evaluation priorities and intended uses of information.
EC (Evalsed, 2009)	• Review of programme documentation. • Analysis of the information system defined in the programme (or related to the programme) and determining the information needs. • Interviewing main stakeholders. • Preparing an analysis of the programmes and theory. • Feedback and review of the above analyses with stakeholders.
USAID (Dunn 2008)	• Verify the Causal Model. • Agree on Purpose of Impact Assessment. • Evaluate Feasibility of Alternative Designs. • Identify Local Evaluation Team.
Leviton (2006)	• Involve intended users of evaluation information. • Clarify the intended program from the perspective of policy makers, managers, and staff and other key stakeholders. • Explore program reality, including the plausibility and measurability of program goals and objectives. • Get agreement on any needed changes in program activities or objectives.

	• Explore alternative evaluation designs. • Get agreement on evaluation priorities and intended uses of information on program performance.
Dawkins (2005)	• Involve stakeholders and intended users. • Clarify program intent (plausibility of goals) and document program as designed. • Determine program implementation. • Work with stakeholders to prioritize key evaluation questions. • Explore designs, measurements, and information systems. • Agree on intended uses.
Wholey (2005)	• Involve intended users of evaluation information in the evaluation planning and design process. • Clarify the intended program. • Explore program reality. • Explore alternative program designs and alternative monitoring and evaluation designs. • Get agreement on monitoring and evaluation priorities and intended uses of evaluation information. • Proceed by successive iterations.
Thurston (2003)	• Bounding the program by identifying goals, objectives, and activities that make up the program. • Reviewing documents. • Modeling resource inputs, intended program activities, intended impacts, and assumed causal links. • Scouting the program or getting a first hand look at how it operates. • Developing an evaluable program model. • Identifying evaluation users and other key stakeholders. • Achieving agreement to proceed on an evaluation.
Smith (1989)	• Determine Purpose, Secure Commitment, and Identify Work Group Members. • Define boundaries of Program to be Studied. • Identify and Analyze Program Documents. • Develop/Clarify Program Theory. • Identify and Interview Stakeholders. • Describe Stakeholder Perceptions of Program. • Identify Stakeholder Needs, Concerns, and Differences in Perceptions. • Determine Plausibility of Program Model. • Draw Conclusions and Make Recommendations. • Plan Specific Steps for Utilization of EA Data.
Rog (1985, in Smith (1989))	• Studying the program's design. • Studying the program's implementation. • Studying the measurement and information system. • Analysing the plausibility of program goals. • Developing different program models. • Determining the uses of information stemming from the planned evaluation.

Annex E: Monk's (2012) tabulation of uses of Evaluability Assessment

Appendix 2: Uses of evaluability assessment in practice

	For which projects?		When?		For what?		
	For major projects	Informally for everything	At the beginning of the project	Just before the evaluation	To clarify/ check program design	To design the evaluation	Decision to evaluate
UNEG	√				√	√	√
UNICEF			√		√		
UNDP				√		√	√
UNODC		√	√		√		√
UN WOMEN	√	√		√	√	√	√
ILO			√	√	√	√	√
USAID				√	√	√	√
JJEC	√	√		√	√	√	
IADB		√	√		√		
	Others (don't use the language of EA, but have similar practices)						
UNESCO				√		√	
Danida				√		√	√
FAO				√		√	
Kellogg Foundation			√	√		√	

Annex F: Example checklists

Anon. 2011. "Using the Evaluability Assessment Tool. Guidance Note 11." **ILO**. http://www.ilo.org/wcmsp5/groups/public/---ed_mas/---eval/documents/publication/wcms_165984.pdf

Dunn, E. 2008. "Planning for Cost Effective Evaluation with Evaluability Assessment". **USAID**. http://pdf.usaid.gov/pdf_docs/PNADN200.pdf

Evalsed, **European Commission**. 2009. "Evaluability Assessment." Source Book: Methods and Techniques. http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed/sourcebooks/method_techniques/structuring_evaluations/evaluability/index_en.htm

Gunnarsson, Charlotte. 2012. "Evaluability Assessment Template". **UNODC.** http://www.unodc.org/documents/evaluation/IEUwebsite/Evaluability_Assessment_Template.pdf

Leonard, Keith, and Amelie Eulenberg. 2012. "Evaluability Brief – Evaluability – Is It Relevant to EBRD?" **European Bank for Reconstruction and Development.** http://www.ebrd.com/downloads/about/evaluation/130305Evaluability.pdf

Monk, Heidi. 2012. "Evaluability Assessment at IDRC". **IDRC**. Not available online. Request via http://www.idrc.ca

Sniukaite, Inga. 2009. "Guidance Note Carrying out an Evaluability Assessment". **UNIFEM.** http://erc.undp.org/unwomen/resources/guidance/Guidance%20Note%20-%20Carrying%20out%20an%20Evaluability%20Assessment.pdf

Related guidance but not including checklists

Reimann, Cordula. 2012. "Evaluability Assessments in Peacebuilding Programming". **CDA**. Not available online. Request via http://www.cdainc.com/cdawww/default.php

Annex G: Outline structure for Terms of Reference for an Evaluability Assessment

This structure has been adapted from the [Better Evaluation website](#) and informed by a reading of Terms of References for Evaluability Assessments produced by five international agencies (UNIFEM, UNFPA, UNICEF, USDoJ, DFID)[12] as well as the 2010 [DFID Evaluation Study Terms of Reference TEMPLATE](#).

1. **Why and for whom the evaluation is being done**

   - Background information about the project, program or policy, including objective, strategy and progress to date
   - Purpose(s) of the Evaluability Assessment, that may include one or more the following:
     - Core purposes – to assess
       - Evaluability of the project design
       - Availability of relevant information and capacity of systems to deliver same
       - The practicality and utility of an evaluation, given the nature of the project and the context in which an evaluation could take place
     - Supplementary purposes
       - To propose refinements to project, program or policy design
       - To propose the development and improvement of M&E systems
       - To propose options for an evaluation design, including
         - Timing
         - Evaluation questions
         - Evaluation methods
         - Resources and expertise
   - Primary intended users and uses of the Evaluability Assessment
   - Key Evaluability Assessment questions
     - See tables on pages 20-23 as a menu offering choices here

2. **How the Evaluability Assessment will be accomplished**

   - Overall scope and approach
     - At what stage in the project cycle will this Evaluability Assessment take place?
       - See page 6 of the main report
     - Is this a mandatory or voluntary Evaluability Assessment, and if the latter, initiated by whom?
     - How is it expected that the results of the Evaluability Assessment will be used?
   - Evaluability Assessment methodology
     - Will this be the first step in an evaluation by the same parties, or an independent prior step that may inform ToRs for a subsequent evaluation by other parties?
     - Will this be a desk review or will evaluators also need contact with project participants in situ?
     - What process steps will be essential?
       - See page 10 of the main report
     - Will checklists be required? If so, with what sort of specifications?
     - What are the risks that need to be considered and managed?
       - See page 24 of the main report

---

[12] Listed in the Bibliography on Evaluability Assessment at http://mande.co.uk/blog/wp-content/uploads/2013/02/Zotero-report.htm

- Evaluability Assessment outputs (See page 22-23 of the main report)
    - Assessment of projects design including recommendations that will make it more evaluable
    - Assessment of data availability and the systems & capacities to make it available, including recommended changes that will improve evaluability
    - Assessment of the context
        - Is an evaluation practically possible? What is needed to make it so (including timing, people & resources required)?
        - Is there demand for an evaluation? Which stakeholder interests can or should be addressed? What evaluation questions and designs could meet their needs, given the project design and likely data availability?

3. Who will undertake the Evaluability Assessment and accountabilities

- Professional qualifications, experience and expertise required for the evaluator or evaluation team
- Roles and responsibilities of the parties, including processes for signing off on the evaluation plan and reports
- Ethics and standards guidelines that may be relevant
- Conflict of interest and eligibility constraints

4. Milestones, deliverables and timelines for the Evaluability Assessment

- What deliverables are required and when – for example, detailed Evaluability Assessment plan, inception report, progress report, interim report, draft final report, final report
    - See page 22, for a list of possible outputs
- Timelines
    - And any associated payment schedule

5. What resources are available to conduct the Evaluability Assessment?

- Budget (if organization's policy allows this to be stated)
- Existing data description, with relevant references in annex if needed
- Key contact persons
- Relevant policies to be referred to

6. Annexes

- Essential background reading to accompany ToRs: References and essential full texts
- Award criteria: How proposals will be assessed, if part of a competitive tender

DEPARTMENT FOR INTERNATIONAL DEVELOPMENT

DFID, the Department for International Development: leading the UK government's fight against world poverty.

Since its creation, DFID has helped more than 250 million people lift themselves from poverty and helped 40 million more children to go to primary school. But there is still much to do.

1.4 billion people still live on less than $1.25 a day. Problems faced by poor countries affect all of us. Britain's fastest growing export markets are in poor countries. Weak government and social exclusion can cause conflict, threatening peace and security around the world. All countries of the world face dangerous climate change together.

DFID works with national and international partners to eliminate global poverty and its causes, as part of the UN 'Millennium Development Goals'. DFID also responds to overseas emergencies.

DFID works from two UK headquarters in London and East Kilbride, and through its network of offices throughout the world.

From 2013 the UK will dedicate 0.7 per cent of our national income to development assistance.

Find us at:
DFID,
1 Palace Street
London SW1E 5HE

And at:
DFID
Abercrombie House
Eaglesham Road
East Kilbride
Glasgow G75 8EA

Tel: +44 (0) 20 7023 0000
Fax: +44 (0) 20 7023 0016
Website: www.dfid.gov.uk
E-mail: enquiry@dfid.gov.uk
Public Enquiry Point: 0845 300 4100
If calling from abroad: +44 1355 84 3132