



WORKING PAPER JANUARY 2022

**ASSESSMENT OF THE EBA'S COUNTRY
EVALUATIONS: QUALITY, USE & LEARNING**

Elliot Stern and Ole Winckler Andersen

Assessment of the EBA's Country Evaluations: Quality, Use & Learning

Elliot Stern and Ole Winckler Andersen

Working Paper January 2022

to

The Expert Group for Aid Studies (EBA)

The EBA Working Paper Series constitutes shorter overviews, surveys, mappings and analyses that have been undertaken to bring about discussion and advance knowledge of a particular topic. Working Papers are not subject to any formal approval process by the Expert Group. Just as in the EBA reports, authors are solely responsible for the content, conclusions and recommendations.

Please refer to the present report as: Stern, Elliot and Winckler Andersen, Ole (2022), *Assessment of the EBA's Country Evaluations: Quality, Use & Learning*, Working Paper January 2022, The Expert Group for Aid Studies (EBA), Sweden.

This report can be downloaded free of charge at www.eba.se

Cover design by Julia Demchenko

Elliot Stern is Emeritus Professor of Evaluation Research at Lancaster University UK and is a Fellow of the UK Academy of Social Sciences. He has worked as a consultant on evaluation design and quality for OECD, the European Commission, CGIAR, World Bank and various UK government departments. He is a past President of the European Evaluation Society and is Editor of the journal *Evaluation: the international journal of theory, research and practice*.

Ole Winckler Andersen has been Head of the Evaluation Department of the Danish Ministry of Foreign Affairs, Deputy Head of the Danish Delegation to OECD and is a former vice-chair of the DAC. He has been a member of the Danish Council for Development Research and a Board member of the European Evaluation Society. He is currently Senior Analyst at the Danish Institute for International Studies, where his research is focused on development finance.

Acknowledgements

We would like to thank many people for facilitating this assessment – but there is only space for a few! We would therefore like to especially thank Lisa Hjelm for managing us with such a light touch; and all her colleagues in the EBA Secretariat, especially Jan Pettersson and Markus Burman for their generosity with their time. Vera Mörner was able to open many Ministry of Foreign Affairs doors and guide us down the right corridors with an unerring sense of where we wanted to go. All those we interviewed were helpful and informative, none more so than the Team Leaders for EBA's Country Evaluations. We however have held our own counsel throughout and responsibility for this report is entirely that of the authors.

Elliot Stern and Ole Winckler Andersen, January 2022

Table of Contents

Foreword by the EBA	1
Summary	2
1 Assessment in a Changing Context	9
1.1 The Scope of this Assessment	9
1.2 Approach	10
1.3 EBA’s evolving context	12
2 Quality, Use and Relevance	16
2.1 Quality criteria and standards	16
2.2 Reviewing the quality of CE reports	18
2.3 Conclusions and implications	35
3 Quality, Use and Learning in Practice	38
3.1 Introduction	38
3.2 Evaluation ‘Quality’	39
3.3 Evaluation use	42
3.4 Learning through CEs	46
3.5 Conclusions and Implications	53
4 Looking Forward	56
4.1 Introduction	56
4.2 An outline agenda for CE strengthening	57
References	60
Appendix 1: EBA Comments and Planned Actions	66

Foreword by the EBA

A considerable part of Sweden's foreign aid is regulated through country strategies. In addition, much aid is allocated to individual countries through other strategies. Still, there are few evaluations of Sweden's engagement in specific countries over longer time periods above individual projects, programmes, or strategies.

Since its inception in 2013, EBA has seen strategic country evaluations (thus not necessarily the same as evaluations of country strategies) as a natural and important, though limited, part of its evaluation portfolio. To date, EBA has published five country evaluations, a sixth is to be published in 2022, and others are in a planning phase.

As part of an endeavour to develop its work in the light of experience gained, in 2020, EBA commissioned Elliot Stern and Ole Winckler Andersen to assess the quality, use, learning and relevance of EBA's country evaluations published so far. The authors have done an impressive job in understanding, digesting, and assessing, not only the individual evaluations, but also many of EBA's methods and procedures. We are sincerely grateful for observations made, which has already led to several, and will lead to further, changes. As such, the report not only impacts EBA-commissioned country evaluations but the portfolio at large.

While the intended use of this assessment is internal to EBA, we do believe that there are many important lessons to learn also for other actors in the development evaluation sphere. We therefore publish the report together with our internal plan for follow-up. This is in line with EBA's remit to actively disseminate lessons learnt, conclusions and recommendations from completed studies, and to operate in an open and transparent manner.

EBA working papers are shorter studies of questions of limited scope or that complements a regular EBA report. Working papers are not subject to a formal decision from the expert group but instead reviewed by the secretariat before publication. The authors are, as with other EBA publications, responsible for the content of the report and its conclusions.

Stockholm, January 2022

Jan Pettersson, Managing Director

Summary

Introduction

This report assesses a portfolio of five country evaluations (CEs) commissioned by EBA between 2015 and 2017 and published between 2016 and 2018. These CEs, in Tanzania, Uganda, Cambodia (the site of 2 studies) and Bosnia-Herzegovina took a long-term view consistent with the long-term priorities and continuities of Swedish development cooperation policies.

The assessment was expected to draw lessons for EBA from these CEs, i.e. to:

- Explore what can be learnt from the evaluations that have been conducted so far, focusing on the use, learning and relevance for Swedish development cooperation.
- Assess the quality of the evaluation reports using the EBA policy for study quality.
- Identify aspects of the EBA country evaluations that could be improved to increase the quality, use, learning and relevance for the Swedish development cooperation?

Because of the focus on learning lessons for EBA in the future, both the five reports and the process of portfolio management and implementation were considered. This included locating CEs in the organisational setting of EBA as well as the institutional setting of Swedish development cooperation.

Changes in Swedish development policies, institutional developments and revised government ‘instructions’ to EBA have also been taken into account to ensure that potential improvements are properly contextualised.

Applying quality standards to CE reports

EBA's recently adopted 'Quality Policy' together with OECD/DAC criteria and standards were applied to the CE portfolio reports. The overarching criteria drawn from EBA's policy included: reliability, usability, and learning; efficiency and implementation; and ethics. The overarching criteria drawn from the OECD/DAC were: relevance, coherence, effectiveness, efficiency, impact and sustainability.

- In terms of **reliability**, evaluation conclusions were generally consistent with evidence collected. The range and specificity of conclusions might have been greater if evaluation teams had had a broader methodological skill-base.
- In terms of **usability** of knowledge produced, CE reports contributed towards usable knowledge but to maximise potential relevance for learning, further synthesis and more effort to translate findings into policy settings would have been needed.
- In terms of **efficiency** and **implementation**, CE teams had to deal with budgetary constraints and problems of information access. As a result, whilst agreed evaluation questions were generally answered, initial ambition was sometimes tempered in the course of implementation.
- In terms of **ethics**, CE teams conformed with standard confidentiality practice, were culturally sensitive and addressed problems when they arose appropriately. More problematic ethical issues such as engaging with partner countries; and adjusting to changing values were not consistently dealt with. Neither EBA nor CE teams employed commonly used evaluation ethical codes.
- In terms of **relevance**, evaluations responded to Swedish priorities but less to country priorities, reflecting a move away from Paris Declaration norms over recent years. Contemporary aspects of relevance related to climate change were not in evidence.
- In terms of **coherence**, CEs were weak in analysing coordination with other donors including multilaterals and the EU and with country plans. In addition, long-term governance – also an aspect of aid coordination – was not addressed.

- In terms of **effectiveness**, CEs were influenced by Paris Declaration thinking even though by the time these CEs were commissioned this was less commonly accepted. In terms of explaining aid effectiveness, CEs addressed relevant material descriptively but not through causal analysis.
- In terms of **efficiency**, this was addressed in various ways, from different perspectives and conceptualisations and with one exception relying mainly on judgement rather than established economic methods and techniques.
- In the DAC formulation of **impact**, i.e. making a difference to the lives of poor people, CEs addressed this extensively relying on detailed description and logical reasoning to reach conclusions. This was more successful applied to poverty reduction than to demonstrating impacts around human rights and democracy.
- **Sustainability** was addressed in CEs in terms of the sustainability of infrastructure, services and policy reforms rather than in terms of environmental sustainability and climate change even though this is an element of Sweden's 'multi-dimensional' poverty reduction approach.

This analysis suggests that EBA's CE portfolio generated a wealth of information that was collected with appropriate care and which supported useful analysis and reasonable conclusions. Although CE contractors selected had strong development cooperation experience, they were less well-acquainted with evaluation methodologies that could support explanation and causal analysis. CE reports, with one exception, therefore relied on descriptive material filtered by judgements based on experience rather than on methods-led analysis.

Limitations of CEs can often be traced to an imbalance between ambition and available budgets. Advance costings by EBA of future CEs together with more integrated planning across EBA's study portfolio could allow for a better balance. Portfolio planning that, for example ensured consistent thematic coverage, would also increase future learning from CEs through synthesis and comparison across CEs.

A consistent weakness of the CE portfolio is the failure to analyse interactions and synergies between Sweden and other donors. This limited the ability of CEs to reliably evaluate the contribution and value-added of Swedish aid. Interaction with other donors is also important in order to better understand aid governance, an area where innovation will be required with Agenda 2030.

Quality, use and learning

Stakeholders and development practitioners involved in setting up and managing EBA's CE portfolio defined quality, use and learning within their own context and in relation to their own backgrounds and expectations. This assessment was generally able to reconstruct these understandings through interviews and documentation reviews, although many key individuals had moved on and there was a striking lack of documentation on key CE portfolio decisions. The assessment also drew on research thinking about pedagogics, knowledge transfer, quality assurance and evaluation use.

EBA has recently agreed on a Quality Policy which once fully operational will act as a quality control system for evaluation/study reports. It is at an earlier stage of developing a quality assurance (QA) process for the entire evaluation cycle. Such a QA system would be able to identify the preconditions or threats to quality prior to the report stage, e.g. in relation to contracting, team expertise, methodological choices, and liaison with partner countries.

The way evaluation 'use' was understood and operationalised spanned the full range of evaluation use 'typologies' in particular instrumental and conceptual uses of evaluation. Some MFA and Sida officials had expected 'instrumental' outputs useful for their short-term work. However, CEs focused on long-term Swedish development policies rather than current contemporary country strategies, sometimes leading to disappointment. Senior staff interviewed were more open to general insights - to 'conceptual' use of evaluations: these reports 'made you think'. Policy decisions will always be influenced by political priorities more than by evaluation findings. Nonetheless potential use or evaluation influence was further limited by the lack of a shared lexicon of what constitutes 'use', 'quality' and 'learning' among those responsible for development cooperation in EBA, MFA, Sida, and in CE evaluation teams.

Longer-term use and usefulness generally depend on organisational memory and the recollections of key individuals. Many staff involved in CE liaison had moved on. We encountered lapses in organisational memory. Staff interviewed including senior officials were not always aware of CE reports coverage and content even when this could have been useful for current development cooperation planning. This raises questions about systems in place in MFA and Sida as well as staff mobility.

EBA would need to decide if it wished to give greater priority to ‘instrumental’ use in future CEs. In the completed CE portfolio this was not prioritised. Nor were all the current ‘steering arrangements’ in place at the time the CEs were commissioned – e.g. Terms of Reference, Reference Groups, ongoing involvement of MFA and Sida. To steer CEs more purposefully would also require the principle of ‘double independence’ as understood by EBA to be reconsidered.

Evaluation use requires that evaluation outputs have to be communicated. EBA’s communication, liaison and coordination strategies with stakeholders were not well developed in 2016–2018. EBA is nowadays better organised to ‘get the message out’. However different communication strategies are suited to different potential uses of evaluation. If the aim is that knowledge and insights continue to be available to inform future thought and action, a more extended and interactive communication strategy would be needed.

In evaluation, learning is not confined to commissioners or stakeholders. Evaluation teams also learn during an evaluation; and not all those who benefit from an evaluation will be known in advance. Evaluation team learning was constrained by CE budgets which led to an over-reliance on documentary sources rather than fieldwork; and limited engagement with in-country actors, including Swedish embassies, partner country governments or other stakeholders.

In pedagogic terms, the dominant mode of learning associated with EBA’s CE portfolio was ‘transmissive’. This mode assumes that giving a learner information will lead to behaviour change – the core task is effective communication. This may work best when what is being ‘transmitted’ is data and facts, rather than, principles or concepts – like democracy and human rights - where contextualisation is required.

In policy settings, the limitations of transmissive approaches to learning are well understood and greater emphasis is placed on collaborative learning, the co-production of knowledge, policy translation, institutionalisation and developing ‘knowledge networks’. Four imagined ‘learning scenarios’ were developed to illustrate different approaches to policy learning and to provide EBA with future options for its own learning and use strategies. They included:

- **Knowledge transfer scenarios**, close to EBA’s default practice at the time the CE portfolio were implemented.
- **Collaborative learning and co-production scenarios**, where stakeholders both in Stockholm and in country would work together on both evidence and dissemination.
- **Policy translation and institutionalisation scenarios**, which would require that problematic or challenging policy choices were examined in their contexts and for different potential users.
- **Knowledge network scenarios**, that would engage relevant policy-shaping communities, e.g. researchers, knowledge brokers, practitioners as well as policy makers in jointly reviewing important but problematic CE findings.

Looking Forward

Although this report supports many conclusions and is suggestive of many recommendations the assessment has avoided making specific recommendations. Instead an ‘outline agenda for CE strengthening’ is put forward, as a way to frame future CE portfolios. EBA drawing on the findings of this assessment may also choose to emphasise other follow-up actions.

The ‘agenda for CE strengthening’ consists of four linked initiatives:

- **Building on EBA’s Quality Policy to develop and embed quality practices in a ‘quality assurance’ system:** This should cover the entire evaluation cycle, and also be supported by EBA’s stakeholders who will necessarily be implicated in QA implementation.
- **Planning future CEs at a portfolio as well as an individual CE level:** This should ensure that all EBA priorities are covered, including choices of country and topic; required team skills are selected; and that synergies are realised between EBA’s country, thematic, methodological and other evaluation studies.
- **Accumulating knowledge across a future CE portfolio to maximise learning:** This would ensure that all evidence about themes (e.g. democracy); methodological innovation; and country engagement could be collected together. This could usefully be supplemented by follow-up assessments ideally conducted in conjunction with MFA and Sida.

- **Developing communications strategies for learning networks:**
This would complement existing communications strategies which are best suited to users/learners interested in responding to known problems. More dialogical communications approaches are needed for dispersed learners located in networks or particular contexts who may become users in the right circumstances.

1 Assessment in a Changing Context

The opening Chapter outlines the scope of this assessment and the approach taken. It also takes note of EBA's changing context. This is important because the context for future Country Evaluations (CEs) will have changed considerably from that in place when existing CEs were planned and implemented.

1.1 The Scope of this Assessment

The invitation to undertake this assessment specified 3 aims:

1. Explore what can be learnt from the evaluations that have been conducted so far, focusing on the use, learning, and relevance for Swedish development cooperation.
2. Assess the quality of the evaluation reports using the EBA policy for study quality.
3. What are the aspects of the EBA country evaluations that could be improved to increase the quality, use, learning, and relevance for the Swedish development cooperation?

Supplementary questions were also posed that like the above aims, emphasise learning from the five evaluations included in EBA's Country Evaluation (CE) portfolio. These evaluations were commissioned in two tranches, beginning in 2015 covering Tanzania and Uganda; and followed up in 2017 including two evaluations covering Cambodia and one covering Bosnia-Herzegovina, published in 2016 and 2018 respectively (Table 1).

Table 1: The EBA Country Evaluation Portfolio

The EBA Country Evaluation Portfolio
• 2016:10 Swedish Development Cooperation with Tanzania – Has It Helped the Poor?
• 2016:09 Exploring Donorship – Internal Factors in Swedish Aid to Uganda
• 2018:10 Nation Building in a Fractured Country: An Evaluation of Swedish Cooperation in Economic Development with Bosnia and Herzegovina 1995 – 2018
• 2019:03 Supporting State Building for Democratisation? A Study of 20 years of Swedish Democracy Aid to Cambodia
• 2019:04 Building on a Foundation Stone: the Long-Term Impacts of a Local Infrastructure and Governance Program in Cambodia

An additional CE has now been commissioned on Ethiopia. However, as this is still only at ‘start-up’ due to Covid-19 delays and current conditions in Ethiopia, it only marginally informed this assessment.

All of these evaluations were expected to take a long-term perspective – as indeed does Swedish development cooperation. This posed the common challenge for these CEs insofar as evaluation as a practice and profession has always tended to emphasise the short term. Although this has started to be questioned as evaluation has extended its reach from projects to programmes and to policies and strategies; taking a whole country perspective over a 20 year plus timescale undoubtedly challenges established evaluation mindsets and toolkits.

1.2 Approach

Our proposal in response to EBA’s invitation tried to balance two polarities both present in EBA’s invitation and implying a somewhat different emphasis. These were:

- The CE Reports as an object of assessment on the one hand and the broader evaluation process which shaped these reports on the other; and,
- A backward look towards the five CEs and a forward look towards potential for improvement in future CEs.

In order to maintain a balance between these polarities, CE Reports were conceived of as dynamically shaped by: CE ‘chronology’ - when studies were commissioned and delivered and how this shaped their ‘opportunity space’ to have influence; the way evaluations are used in Swedish development cooperation given the mechanics of country strategy-formation; and how evaluation ‘processes’ are managed by EBA and its partners. This framework built on EBA’s aims and questions. All of the interactions, sketched out in Figure 1 below, are also embedded in an international ‘state of the art’: what is known about how evaluations and assessments of development evaluations is practiced internationally.

Figure 1: Framework



Source: The authors.

Broadly speaking, this framework was followed and was useful, although inevitably we have become aware of further complexity given EBA’s independence; positioning vis a vis MFA and Sida; and the way strategy making occurs in Sweden’s distinctive and dispersed aid management system.

Undertaking this assessment highlighted the diversity and interactions around these five CEs, which effectively began to ‘pilot’ some of the alternative ways that any future CE portfolio could be implemented. This assignment therefore provides an important accompaniment to any pilot scheme: an opportunity to reflect and learn.

The limited evidence base together with the kinds of questions EBA posed – about learning, quality, strategy formation and the evaluation process – made a qualitative approach inevitable and appropriate. On closer inspection, this portfolio is also diverse in different ways. For example, the Terms of Reference were not standard; and those proposing CEs all exercised discretion in study design and implementation.

This diversity made a straightforward comparative approach difficult: tabular comparisons of ‘quality’ and ‘use’ across cases would have had to be constantly explained and qualified, e.g. in terms of different assumptions made in ToRs and different decisions made by Team Leaders and stakeholders during the course of each CE in the face of different challenges.

This has ethical as well as analytical implications. We did not think it reasonable to assess the quality of individual reports that were not expected to follow common standards – EBA’s Quality Policy did not then exist – and which furthermore have already been ‘quality assured’ by EBA prior to publication. Rather than assessing individual CEs, we have concentrated on the different ways the quality of the entire CE portfolio can be understood. Because of the way CEs have been shaped by the wider EBA and Swedish aid system, we attempted both a ‘summative’ assessment of the portfolio and an examination of how key concepts such as quality, use, and relevance were defined in practice and what accounts for these definitions.

The assessment relied on:

- Interviews/discussions with key informants, including all Team Leaders of these CEs; and those from EBA, MFA and Sida involved in these CEs at the time, or in similar roles today, given the inevitable career mobility over the last five years.
- A review of a considerable dossier of documents¹, both directly related to each CE (i.e. invitations, proposals, reports, feedback from EBA); policy, practice and guidance reports from EBA, MFA and Sida; and other literature and documentation related to evaluation in general, and the evaluation of international development in particular.

1.3 EBA’s evolving context

As already noted, this exercise both looks backwards to the completed CE portfolio and forwards to strengthening future CEs. Things have not remained static since the CE portfolio was commissioned. EBA’s operating environment and practices have continued to evolve, often in ways that have implications for CEs as well as other parts of EBA’s study portfolio. These changing practices are also informed by an evolving understanding of about how to communicate and coordinate with policy actors; and how to adjust to the changing international world of Agenda 2030.

¹ Interviews and documents are listed in the References to this report. Interviewees also provided us with a variety of internal notes and memos. These have been used but are not cited as source material.

This chapter addresses these topics in order to set the scene for following chapters which assess the quality of the existing CE portfolio; describes the ways CE stakeholders understood quality, use, and learning; and discusses possible ways forward.

EBA is a relatively young organisation that has been evolving and developing together with its operating environment since its launch in 2013. Changes in EBA's practices are inevitably influenced by contextual changes, many of these coterminous with, or subsequent to, the implementation of the CE portfolio. Significant 'contextual' factors include:

- Swedish government priorities for aid and the way it should be delivered changed with changes in governments².
- New set of government 'instructions' were issued to EBA, re-emphasising the importance of evaluations in 2016 (see excerpt next page).
- A further organisational review by Statskontoret (the public management agency) took place in 2018 acknowledging ongoing improvements and emphasising what was still needed.
- A new Sida central evaluation function also created in 2018, which has since then progressively clarified its role and functions and has the potential to impinge on EBA's work.

In parallel with these contextual changes, EBA's practices have also evolved:

- Policies and procedures regarding operational planning, commissioning, and managing studies and evaluations continued to evolve and informal practice became formalised.
- The volume of EBA studies commissioned and published together with seminars organised increased – CEs as currently conceived are a very small part of overall EBA activity.
- EBA's Communications policy was revised and resources devoted to disseminating studies increased.

² See: Policy framework for Swedish development cooperation and humanitarian assistance Government Communication 2016/17:60 and Guidelines for strategies in Swedish development cooperation and humanitarian assistance Annex to Government Decision 21 December 2017 (UD2017/21053/IU).

- An EBA Quality Policy was drafted, and finally adopted in 2020.
- Consultation and coordination between EBA and MFA and Sida have increased in scope and scale through both formal and informal channels.

These changes have modified the landscape for all EBA evaluations and studies as well as for CEs. How future CEs will be commissioned, supported, and interfaced with MFA and Sida; and how their reports disseminated will differ in many ways from the existing CE portfolio.

One important part of the CE evaluation landscape is shared with Sida. We were at first convinced that there was a natural division of labour between EBA and Sida. For example, EBA is committed to the ‘long term’ and Sida’s focus is on reinforcing shorter term strategy cycles; and EBA’s country and thematic focus sits alongside an evaluation strategy in Sida that emphasises projects and programmes rather than portfolios. However, having become familiar with Sida’s plans, we became more aware of potential overlaps between EBA and Sida – even though the 2018 Statskontoret report downplayed such risks. Two examples of this: the enhanced independence of Sida’s evaluation unit and its planned increases in ‘strategic’ evaluations; and EBA’s Operational Strategy (2019) that indicates analyses of ‘...questions relating to the relevance of aid in the short, medium and long term’. We see it as inevitable that EBA’s future ‘opportunity space’ will be affected by Sida’s plans and vice versa. This in part explains the intensification of consultation between EBA and Sida over the last five years.

Changes subsequent to the delivery of EBA’s CE portfolio between 2016 and 2018 and the present; and the direction of travel of changes now evident, highlight the dangers of simplistically transposing assessment findings from the existing CE portfolio to future CE ‘improvements’. We have therefore tried as far as possible to take EBA’s changing context into account as part of this assessment.

When discussing the existing CE portfolio, we consider the ‘opportunity space’ available to these evaluations to inform development cooperation policies given their content and timing. Here it is worth noting that Sweden’s current development cooperation policy documents appear to place relatively little emphasis on the importance of evaluation or research evidence to inform policy and strategy development, although this is

briefly mentioned in the 2017 Strategy Guidelines. It appears from an outsider perspective, that it is only recently and especially since 2018 that the post SADEV evaluation system has truly started to take shape.

Nonetheless a move towards a more knowledge-intensive or evidence-based approach to development assistance was given added impetus by the 2016 Instructions (Committee Directive 2016:71), which emphasised that ‘well-founded decisions require knowledge’ (Table 2).

Table 2: EBA’s instruction (excerpt)

The expert group shall
<ul style="list-style-type: none">• order or carry out evaluations, analyses and studies of the implementation, results and effectiveness of development assistance;• independently set guidelines and methods for the activities, prioritize study areas and themes, decide on which evaluations, analyses and studies are to be carried out and decide on publication;• ensure the quality of the analysis and evaluation activities, whereby the OECD-DAC's principles and quality standards for aid evaluation should provide guidance where appropriate;• actively disseminate lessons, conclusions and recommendations from completed studies to relevant target groups in an appropriate manner.
<p>The expert group shall strengthen the overall analysis and evaluation activities in the area by complementing other such activities and focus in particular on issues and study areas where its independent position has a clear added value.</p>
<p>The studies must address issues that are strategically important for development assistance and development policy.</p>

Much of the content of these instructions sets the directions in which EBA has been moving, for example with regard to reinforcing quality of its work, and dissemination of studies to relevant target groups.

2 Quality, Use and Relevance

2.1 Quality criteria and standards

A core task of this exercise is to assess the quality of the CE portfolio. This section assesses the quality of CE Reports. A distinction has been made between an assessment of the final product, a report, and an assessment and analysis of the processes that shaped these reports. In the following section, the CE portfolio is assessed in a qualitative and discursive ways so as to better understand what ‘quality’, ‘use’, and ‘learning’ means in the EBA and CE context.

How to frame and exercise judgements about evaluation quality has preoccupied the evaluation community and those managing public policies for 40 years or more. Unsurprisingly therefore a number of different ‘frameworks’, ‘toolkits’ and ‘guides’ have been developed to support such assessments. Whilst agencies often have their own frameworks customised to their circumstances most stem from one of two main sources: the AEA ‘Joint Standards’ itself originally rooted in educational evaluation but over time broadened to ‘programme evaluation’ more generally; and the OECD/DAC Evaluation Criteria and Quality Standards, rooted in international development cooperation.

All quality frameworks are made up of a number of ‘building blocks’:

- An assessment of the methodology: i.e. the appropriateness, reliability, and strength of a chosen approach, covering in particular theory, methods and data analysis.
- An assessment of the evaluation ‘object’: i.e. the criteria best-suited to identifying and evaluating the attributes of a programme, policy, or intervention.
- An assessment of evaluator performance and capability: i.e. whether an evaluator is knowledgeable, competent, independent, free from bias, and behaves ethically.
- An assessment of report quality: i.e. whether a report is well structured, communicates effectively and relating to what was proposed, and to what has been learned.

These building-blocks can be variously combined or packaged but most approaches to quality in evaluation emphasise some building blocks more than others.

- EBA's Quality Policy is built around four principles – Reliability, Use, Effectiveness and Ethics, mainly drawing on American Evaluation Association 'Joint Standards', but also overlapping to an extent with OECD/DAC 'Quality Standards' (see below).
- The OECD/DAC 'Evaluation Criteria' focus mainly on aspects of interventions – relevance, effectiveness, efficiency, impact, and sustainability – coherence was added in the most recent revision (December 2019).
- The OECD/DAC 'Quality Standards' are a guide to good practice in the design, implementation and reporting of development evaluations.

The EBA Invitation for this exercise indicated that the CEs should be assessed using the EBA's own recently agreed Quality Policy. This policy is strongest when assessing the strengths of an evaluation report, focusing on how well it was conducted and the quality of the final report. In our judgement, relying only on the EBA policy would not sufficiently cover intervention content alongside report adequacy. We therefore decided to apply both EBA's Quality Policy and the OECD/DAC Evaluation Criteria to the CE portfolio. Although there is some overlap between these two quality approaches, they are mainly complementary.

The overarching criteria of EBA's Quality Policy as we have operationalised them, are:

- **Reliability**, which concerns the soundness of theory, methods, and analysis; whether an evaluation will deliver what it promised; whether data and analysis is sufficient and supports conclusions.
- **Usability and learning**, which concerns whether the report, and especially its conclusions, are clear and accessible; whether stakeholder engagement and dialogue was encouraged; and whether new knowledge has been produced.
- **Efficiency and implementation**, which concerns the reasonableness of the evaluation budget; and how well an evaluation has been implemented.
- **Ethics**, which concern protecting sources, cultural sensitivity, and ethical awareness; whether ethics is mentioned; stakeholder dialogue and consultation.

The overarching criteria of the DAC's Evaluation Criteria, as we have operationalised them, are:

- **Relevance:** Is the intervention doing the right thing: responding to country and global needs and Swedish priorities?
- **Coherence:** Is there coordination with other donors? Is there coherence with country plans and priorities?
- **Effectiveness:** Is the effectiveness of Swedish aid considered and assessed?
- **Efficiency:** Is efficiency/value for money, considered in programmes implemented?
- **Impact:** Is Sweden's programme likely to make a difference for poor people or for countries and their governments?
- **Sustainability:** Is the continuation and viability of interventions discussed?

2.2 Reviewing the quality of CE reports

As will be evident, this is intended as an overview rather than a forensic examination. There has been no attempt to review all 60 items in EBA's Quality Policy's 'Form for Assessment'; nor all the sub-items in the DAC Evaluation Criteria. Rather, these two approaches have been used to decide what to prioritise when assessing CEs, and in particular the CE Reports. Although this was built up from a case-by-case assessment, our purpose was not to assess individual CEs. The ethical risks of retrospectively applying criteria and standards not in place at the time these evaluations were commissioned has already been noted. The intention rather is to provide a quality overview at portfolio level.

What follows, begins by summarising in a box, key points, and then expands on each quality criterion. In the first instance, we apply a streamlined version of EBA's Quality 'Assessment Form'; and follow this by applying DAC Evaluation Criteria. The section mainly applies criteria to the reports but on occasions also draws as background, on interviews and documentary analysis in order to make sense of what can be read into CE reports. Issues identified here are also further discussed in the next Chapter which considers the practice and dynamics of 'quality', 'use', and 'learning' in greater detail.

Each criterion is given a summary ‘assessment statement’ at the end of each sub-section.

Table 3: Criterion EBA 1 – Reliability

Criteria EBA 1	Key points
<p>Reliability: Soundness of theory, methods, and analysis? Likely to deliver what was promised? Is data sufficient? Does analysis support conclusions?</p>	<ul style="list-style-type: none"> • Strength of evaluation in short and long term • Macro theory and explanatory theory • Methodological approaches • Non-representation of evaluation community • Problems of data access

Evaluation has traditionally not been strong assessing the long-term. In the social sciences more generally, causal analysis is most capable when dealing with short time-scales and/or well-defined ‘objects’ of evaluation. The one case in the CE portfolio that applied strong analytical techniques did so in relation to a well-defined programme rather than an entire country aid package.

CEs have been pragmatic drawing on and adapting existing theories and methods. Theory is especially important in a setting like CEs in order to explain causes as well as account for results. At a high level of analysis CEs used macro theories to frame their analysis (e.g. macro-economic explanations of poverty reduction; normative theories of aid effectiveness like the Paris Declaration; and theories of Political Economy and Democracy). These macro theories are useful when focusing a study although they usually have a value component and rely on a shared consensus about these values. At an intermediate level, many CEs accepted the *logic* of analytical methods but applied them as *principles* rather than as techniques. Examples of this included ‘counterfactual thinking’, and in principle, adaptations of Contribution Analysis and Process Tracing approaches. The use of ‘critical junctures’ as a way to capture changes over time was more operational and proved useful. Such frameworks were valuable in structuring combinations of quantitative and qualitative data. The resulting analyses relied heavily on the judgement of teams that were highly experienced and better able than most to reach reliable judgements about causal processes. However, CEs also sometimes aspired to greater rigour and would have benefited from more methodological expertise in evaluation teams for at least part of their workplans.

The teams selected for these CEs with one exception were practice led rather than methodology led. Several CEs had weak descriptions and discussions of their methodological approach. It is noteworthy that the mainstream evaluation community was not represented in the CE portfolio and as a result, contemporary evaluation approaches were also under-represented. For example, Theory Based Evaluation (TBE) approaches to impact evaluation have evolved specifically to address complex settings where the interventions of interest cannot be isolated from other potential causal factors, and where the isolation of causal mechanisms in context are especially important. In plain language, because country programmes and strategies over the long-term are not amenable to simple cause/effect analysis and country contexts make a difference, additional methods are needed to answer the kinds of questions EBA posed for CEs.³

There are of course arguments in favour of contracting non-standard teams for CEs, so as to encourage ‘out-of-the-box’ thinking. On the other hand, there are also costs of not including in these efforts, stronger methodological expertise generally, and those familiar with state-of-the-art TBE methods in particular.

At the same time, the portfolio appears to have improved on earlier generations of CEs. One familiar challenge of CEs, case-study selection and how to link cases with country-wide generalisations, was generally well-designed and justified – a step forward from the generation of CEs discussed in the OECD Workshop on Country Evaluations in 1999.

As was inevitable, access to data proved a problem for most CEs. There were instances in this portfolio when analytic choices were not those preferred by CE authors. Rather the choice was determined by data availability – for example by focusing on income poverty rather than multi-dimensional poverty. CEs were inventive in using proxy and secondary data, and not only in the well-known instance where geospatial lighting data was used. Previous evaluations, historical records, and repurposed administrative data were also used creatively. Although previous evaluations were used when available, there were concerns reported about the usefulness of Sida evaluations during certain periods when Sida relied mostly on ‘results’ reporting. One cross-portfolio opportunity for learning would be to collate ‘good practice’ examples of data sourcing and data analysis to inform future CEs.

³ See for example Cartwright N (2007), Stern (2015) for fuller discussion of methodological debates.

*In terms of **reliability**, CEs can be regarded as reliable within certain parameters (i.e. assumptions shared within the policy community). The experience of CE teams was sufficient to give confidence in their judgements when dealing with incomplete data and sometimes confusing evidence. Conclusions were generally consistent with evidence but at a high level of generalisation and contextually qualified – appropriate for country-level evaluations. The range and specificity of conclusions might well have been greater if the teams had had a broader methodological skill-base.*

Table 4: Criterion EBA 2 – Usability and learning

Criteria EBA 2	Key points
<p>Usability and learning: Is the report clear and accessible? Are conclusions clear? Was stakeholder engagement and dialogue encouraged? Has new knowledge been produced?</p>	<ul style="list-style-type: none"> • Evaluation rather than practitioner or policy reports • Managing the interface with stakeholders • Reliable knowledge accumulates across evaluations • Handling critical conclusions

These were evaluation reports. They were authored by teams that included development practitioners but were not drafted for a policy readership, even though they are all reasonably well structured and written, have summaries and include ‘lessons learned’ sections. Evaluation use usually requires a process of mediation and translation in order to communicate effectively with policy practitioners or other ‘intended’ users. Mediation begins early in evaluation cycle with stakeholder consultation and involvement. Limited early and continuing involvement of stakeholders may explain some weaknesses of CE reports in terms of their usability. However, one of the challenges of assessing usability of CEs is that the process of commissioning CEs did not favour reports likely to feed directly into decisions or planning. Instrumental use in Michael Patton’s sense of ‘intended use by intended users’ is the exception not the rule in this portfolio. (These topics are discussed further in the following Chapter.)

It is long established among evaluators that ‘learning’ does not come from a single report. Individual ‘bits’ of information, analysis and even insight need to be cross-checked and accumulated before they can be treated with confidence. Looking across the CE portfolio there are opportunities to synthesise the content of reports so as to begin this process of accumulation. For EBA to gain the maximum from what has been

invested in CEs so far, would require a systematic identification of what can begin to be accumulated. This could for example include findings on substantive topics such as capacity-building and sustainability; or strategies when working in difficult contexts for democracy and human rights. There are also opportunities to accumulate methodologies applicable (or not) to country-level evaluations, such as approaches to framing and theorising; or data sources and analysis in different settings. As words like ‘context’ and ‘setting’ imply, knowledge is contingent rather than universal. One of the main aims of accumulation is to map out these contingencies: to begin to clarify what holds true under what circumstances.

The content of CE reports can be challenging and critical to MFA and Sida. The openness of administrations to direct critical feedback varies and some reactions to CE reports indicate that lines were crossed in some cases. Although these reactions can be moderated by consultation and dialogue – and by the way critical material is worded – providing unbiased feedback is the distinctive strength of independent evaluations. How expectations are managed is however important. In one case the negative reaction of key stakeholders interviewed seemed to stem from an expectation that they would benefit from direct useful inputs although the evaluation in questions was never designed with this in mind.

CEs are reasonably accessible in terms of clarity but would need mediation and translation into policy/practice settings to maximise their usability. Mediation includes well-tried approaches such as preparing short mini-reports or ‘policy briefs; running workshops for anticipated target-groups; and relying on those in ‘translation’ roles such as a research, policy or evaluation unit in a ministry or agency, able to match lessons to their own context. EBA’s communications strategy already supports some but not all of these activities.

*In terms of **usability** of knowledge produced, CE reports can be said to contribute towards usable knowledge but to maximise potential relevance for **learning**, further synthesis and more effort to translate findings into policy settings would have been needed.*

Table 5: Criterion EBA 3 – Efficiency and implementation

Criteria EBA 3	Key points
Efficiency & implementation: Is budget reasonable? how well was evaluation implemented?	<ul style="list-style-type: none">• Project and wider systemic efficiencies• Access to stakeholders• Budgets and cost estimates• Enabling and proactive management• Delivering as promised

CE teams include experienced practitioners making it a reasonable assumption that they understood and applied basic principles of project management. We know from interviews with team leaders and MFA/Sida counterparts that there were efficiency and implementation problems at a broader level. Access to stakeholders in MFA and Sida was not always easy, and access to case study material and country informants was often difficult. Some difficulties are inevitable in long-term evaluations as key individuals change jobs, records are lost, and people forget. It appears that EBA adapted and learned quite early on the importance of acting as a gatekeeper/intermediary but even so, access was never straightforward. It does not appear that partner countries were consulted or involved in CE approval and planning. This may also have contributed to access and efficiency problems.

The low level of CE budgets was a serious impediment for the first two CEs commissioned and continued to be a constraint even when budgets for subsequent CEs was moderately increased. (The problem does appear to have been recognised given the further, more significant increase in budget for the recently commissioned Ethiopia CE.) Budgets limited the scale of activities and CE ability to collect primary data.

This aspect of efficiency often depends on a pre-commissioning estimation of likely costs given the evaluation questions being asked and the activities needed to answer them. Estimates also have to reflect the strategic importance of an evaluation. It does not appear that pre-commissioning estimation of this kind took place.

Team leaders described implementation management by EBA managers and by Reference Groups as enabling and supportive. How far any commissioning body stands back, supports or actively manages an evaluation portfolio is a matter of choice: any approach has its strengths and weaknesses. There are however circumstances where active

intervention might be expected. Because of a crisis in one embassy there was no possibility of one CE undertaking its work in country. As the proposed evaluation approach required both an analysis of Sweden’s policies and their consequences this resulted in a limited and unbalanced report. Our interviews suggested that although initially it was accepted in MFA and Sida that access problems were not within the control of the study team, later negative judgements about ‘report quality’ appeared to ignore why this was. More proactive management may have led to greater efficiency in this case; and at the same time mitigate reputational risk for EBA.

The extent to which evaluations deliver on what they proposed is a commonly used measure of efficiency. Although CEs generally answered evaluation questions, they did not deliver on their more ambitious promises to develop or validate a more generalisable model for country evaluations. These promises were probably over-optimistic. However, it was noteworthy that neither study teams nor Reference Groups appeared to have been concerned about this discrepancy between proposals and outputs.

*In terms of **efficiency and implementation**, the experience of teams countered most inefficiencies which stemmed from budgetary constraints and ‘systemic’ problems such as information access. While agreed evaluation questions were generally answered, initial ambition was sometimes tempered in the course of implementation.*

Table 6: Criterion EBA 4 – Ethics

Criteria EBA 4	Key points
Ethics: protects sources, cultural sensitivity and ethical awareness, is ethics mentioned; stakeholder dialogue and consultation	<ul style="list-style-type: none"> ● Evaluator ethics ● Cultural sensitivity ● Ethical codes ● The ethics of aid interventions ● Involving stakeholders ● Involving countries and Agenda 2030

Ethical considerations were not referred to in ToRs for these evaluations. The importance of ethics in evaluation practice is in part comparable to the ethical obligations of other practitioners and applied researchers. The evaluator must ‘do no harm’. In evaluation it is especially important to protect the confidentiality of sources. Informants share confidential

information and their anonymity should be protected. Although in some reports it is possible with some effort to identify individuals or at least role-holders, it appears that this aspect of ethical behaviour was well-attended to.

Another aspect of ethics is 'cultural sensitivity'. This relates first to the behaviour of evaluators when working in other countries and cultures. CE teams were experienced internationally, indeed some had direct experience of the countries in which these evaluations took place. It appears that teams were appropriately sensitive and in some cases, this was supplemented by local support staff who were able to act as interlocutors. In one case ethical considerations were raised by stakeholders regarding the independence and impartiality of some teams with prior connections in-country. These considerations were addressed explicitly and appropriately in the particular CE report. However no 'ethical codes' were in place for teams or EBA managers to refer to in cases such as this. Being explicit about ethical expectations in advance is common nowadays in evaluation practice. For example, in international evaluations, adherence to UNEG ethical guidelines⁴ is often a contractual requirement. Furthermore, there was no standard requirement for reports to report on ethical dilemmas and decisions.

A more fundamental ethical matter is the ethics of aid itself: i.e. the ethical consequences of aid disbursements. In a long-term evaluation it is easier to take a detached view of the benefit or harm done by aid. This is complicated because of changing values and ideologies. Some policies vigorously supported in the 1970s might be judged very differently in the early 21st century. Indeed, it could be argued that the lens of the evaluator looking backwards were themselves time bound given that commitment to the Washington consensus; and adherence to Paris Declaration principles is less now than when these reports were written. In these CEs, the consequences of aid volatility and various Swedish policy 'pivots' usually associated with changes in government; and how to reconcile value and political differences between partner countries and Sweden e.g. over human rights would fall into this ethical category. Ethics through the rear-view mirror merits further discussion

Overall, there was only limited input from countries into individual CEs. It is not clear that Countries were consistently consulted when the portfolio was designed or commissioned although national authorities

⁴ <http://www.unevaluation.org/document/detail/102>

were usually informed once an evaluation was underway. Such consultation was mainly in support of fieldwork and in-country data collection, e.g. accessing administrative data. Feedback to country governments sometimes happened as part of post report communications. However, conclusions and ‘lessons learned’ in reports were directed towards Swedish, not country use. Arguably, Agenda 2030 makes it likely that involvement of countries and other beneficiaries in evaluation will become a stronger expectation in future.

*In terms of **ethics**, CE teams conformed with standard confidentiality practice, were culturally sensitive, and addressed problems when they arose appropriately. More problematic ethical issues such engaging with partner countries; and adjusting to changing values through the ‘rear view mirror’ were not consistently dealt with. Neither EBA nor CE teams relied on evaluation ethical codes that are in wide circulation.*

Table 7: Criterion DAC 1 – Relevance

Criteria DAC 1	Key points
Relevance: Is the intervention doing the right thing: responding to country and global needs and Swedish priorities?	<ul style="list-style-type: none"> ● Aid strategies respond to both Country and Swedish priorities ● Tensions between democracy/human rights and government priorities ● Agile response to such conflicts but in extremis Swedish priorities carry most weight ● CEs document changing priorities around ‘Ownership’ & Budget Support ● Global goals associated with Agenda 2030 less evident in long term CEs

As described in country strategies and in CE evaluations, interventions supported by Swedish aid were chosen because they were relevant to country circumstances; and were significant aid investments for Sweden. Priorities such as poverty reduction, post conflict reconstruction and helping develop basic infrastructures and governance capacities, can easily be justified by objective descriptions of country circumstances. However, Swedish aid has the added dimension of support for democracy and human rights and arguably a stronger than average aversion to corruption. The CEs describe how these normative priorities are interwoven with economic and capacity building programmes. The country strategies therefore respond to Swedish, as well as country priorities. However, these two sets of priorities have sometimes come into conflict and the CEs describe how in those circumstances Sweden’s ‘needs’ are decisive.

The most clear-cut example of this was the decision to cease the Cambodia country programme even though the two CEs focussed on Cambodia demonstrated considerable success; including agile programme redirections to reconcile conflicting goals.

A cross-cutting theme that was strongly present in the CE’s and that reflects the period spanned by these evaluations, is adherence to Paris Declaration Principles, and in particular to Country Ownership as the rationale for donor preference for Budget Support. Not only was there a turning away from Budget Support because of fears of corruption by Sweden and other donors, but the debate about country ownership highlighted how different notions of ‘doing the right thing’ sometimes contradicted each other. For example, CEs highlighted arguments within the aid community about the advantages and disadvantages of directing funds through civil society channels rather than through national governments. This dispute even raises questions about the impartiality of some of these CEs in this respect. Some CEs treated the relationship between country ownership, Budget Support, and aid effectiveness as an empirical as well as normative proposition. On the other hand, the analysis presented was mainly descriptive and suggestive rather than demonstrating these links.

Relevance in DAC Evaluation Criteria includes ‘global’ needs, reflecting the shifts towards SDGs and Agenda 2030. Global goals are present in the interventions evaluated by CEs (poverty reduction, human rights etc) but with less focus on climate change which does not feature directly in the programmes evaluated.

*In terms of **relevance**, evaluations responded to Swedish priorities but less to country priorities, reflecting a move away from Paris Declaration norms over recent years. Contemporary aspects of relevance related to climate change were not in evidence.*

Table 8: Criterion DAC 2 – Coherence

Criteria DAC 2	Key points
<p>Coherence: Coordinated with other donors?</p> <p>Coherence with country plans and priorities?</p>	<ul style="list-style-type: none"> • Coordination with other donors and governance noted but not analysed • EBA posed ‘contribution of aid’ questions that required an examination of Sweden’s value-added • CEs chose to examine the contribution of Swedish aid alone

Criteria DAC 2	Key points
	<ul style="list-style-type: none"> • Coordination and value added especially important as part of regional strategies (e.g. with EU) and where multilateral donors present in country

Although CEs were representative of important Swedish investments there were always other donors also operating in these countries. The evaluations identify other donors – USAID, Switzerland, EU, UNDP, World Bank and DFID – and noted that Sweden worked together with these donors in-country. The weaknesses of cooperation with a particular donor may be noted and the strength of ‘harmonisation’ between donors is also noted. However, there was no analysis of coordination arrangements; or of the value-added of Swedish aid in combination with other donor programmes.

These are important gaps in coverage, first because Statskontoret identified ‘long-term governance’ as one of reasons for undertaking CEs. Second, the evaluation questions posed by EBA anticipated interaction between donors. To evaluate whether ‘Swedish aid *contributed* to poverty reduction’ (1st tranche CEs) or ‘*contributed*’ to the achievement of specified development objectives such as democracy (2nd tranche CEs) inevitably sets Swedish aid into the overall mix of other donor investments and objectives. These questions were understood instead as requiring evidence that positive outcomes were associated with Swedish aid inputs.

The centrality of coherence was even more important when in one CE it was noted that Sweden was part of a regional EU programme. More generally, the proportion of Swedish aid that is channelled through multilateral programmes makes it likely that interactions, value-added and coordination were relevant in all CEs.

Coherence with country plans and priorities has already been discussed in relation to ‘Relevance’ and ‘Ethics’. The missing piece of the puzzle that would speak to ‘Coherence’ as a criterion would have been an assessment of governance arrangements. Nonetheless some of the challenges for Sweden managing conflict and coherence with country governments, decentralised authorities and other public agencies has been documented in these CEs. This also is a topic that merits synthesis across CEs in order to maximise what has been learned on the subject.

In terms of coherence, CEs were weak in analysing coordination with other donors including multilaterals and the EU and with country plans. In addition, long-term governance – also an aspect of aid coordination – was not addressed despite being signposted by Statskontoret.

Table 9: Criterion DAC 3 – Effectiveness

Criteria DAC 3	Key points
Effectiveness: Is the effectiveness of Swedish aid considered and assessed?	<ul style="list-style-type: none"> • Effectiveness identified with Paris Declaration agenda • Effectiveness of specific Swedish initiatives also considered • Different phases of Swedish aid and its effectiveness are identified • These phases identify factors likely to influence effectiveness • However, analysis does not integrate contextual factors-contribution of Swedish aid therefore difficult to separate out

During the period in question, aid effectiveness was often discussed as it is in all but one CE in this portfolio. It is identified with the tautologies of the Paris Declaration: if the PD principles are adhered to aid is by definition effective. Nonetheless CEs are also critical of this position and identify aid effectiveness with development outcomes such as poverty reduction, education participation, agriculture as well as democracy and institutional or market strengthening.

Because of the extended period covered in these evaluations, CEs distinguish between stages or phases and differentiate the effectiveness of aid depending on Swedish priorities and policies, country circumstances, and relationships with country governments. The discussions of these stages describe many of the contextual factors that shape aid effectiveness at different times. However, these contextual factors are not integrated into structured analysis.

CEs combine a top-down and bottom-up perspective and many differences in aid effectiveness are linked to changes in Swedish policy, modes of planning and management by Sida; and dialogue and implementation arrangements.

CEs can point to successes and failures of initiatives, projects and sectors in which Sweden has directed significant proportions of its aid budget. However, the extent to which these outcomes are ‘caused’ or contributed

to by Swedish aid is less clear. What in evaluation language is described as a contribution rather than an attribution perspective, would attempt to disentangle Swedish inputs from other donor inputs, secular trends, government policies etc. CEs do sometimes use the language of ‘contribution’, but they do not employ methods that would be needed to demonstrate causal and contributory pathways.

*In terms of **effectiveness**, CEs were influenced by Paris Declaration thinking even though by the time these CEs were commissioned this was less commonly accepted. In terms of explaining aid effectiveness, CEs addressed relevant material descriptively but not through causal analysis.*

Table 10: Criterion DAC 4 – Efficiency

Criteria DAC 4	Key points
Efficiency: Is efficiency/value for money, considered in programmes implemented?	<ul style="list-style-type: none"> • Efficiency is not referred to in CE ToRs • Efficiency also not central in CEs but other related terms are referred to • Problems applying efficiency to priorities such as democracy, human rights & gender • Two CEs address efficiency directly: using a micro-economic and a value-for-money approach • Reference made to budget support efficiencies by reducing transaction costs

Efficiency does not appear in the ToRs/Invitations for CEs even though the word appears frequently in policy documents and country agreements alongside other rhetorical terms such as transparency and accountability. Unsurprisingly therefore ‘efficiency’ is not a common theme across CE reports. Nonetheless various expressions of efficiency do appear in CE reports. For example, reference is made to cost-effectiveness, the dangers of aid proliferation into multiple small projects, varying aid volumes and the risks of corruption. It was noted by one CE that it is difficult to apply notions of efficiency to support for democracy and human rights. One CE did undertake a value for money assessment of gender equality.

Two CEs do address efficiency considerations directly:

- One takes a rigorous micro-economic approach to ‘costs per unit of output’ and identifies project impacts in relation to costs. This is the only CE that applies a recognised methodology to efficiency analysis. This reflected both the disciplinary specialisation of the evaluation team; and the kind of intervention that they focussed on.
- Another CE uses ‘value-for-money’ as a core concept and distinguished between proportions of the Swedish portfolio in country that was good or poor value for money. Report authors recognise that their conclusions rely on ‘informed judgements’ by evaluators although these judgements were made on sound basis, e.g. Sweden’s contribution to sector; overall performance of that sector, and timeliness.

During the period when Paris Declaration was being prioritised, it was argued that Budget Support was efficient on the grounds that this reduced transaction costs.

*In terms of **efficiency**, this was not required in ToRs but was addressed in various ways, from different perspectives and conceptualisations and with one exception relying mainly on judgement rather than established economic methods and techniques.*

Table 11 – Criterion DAC 5: Impact

Criteria DAC 5	Key points
<p>Impact: Is Sweden’s programme likely to make a difference for poor people or for country government?</p>	<ul style="list-style-type: none"> • Quality criteria related to impact were not designed for country programmes • Traditional evaluation concepts and methods also do not match the CE context • CEs rely for their evaluation of impacts for poor people on assembling existing evidence, applying logical arguments and reaching expert judgements • In general CEs report qualified success in Sweden’s effects on the lives of poor people • Impacts on governments are described in detail and should be reviewed and synthesised across the CE portfolio

Impact is a hotly debated word in evaluation circles. It commonly means:

- Final results rather than outputs or intermediate results, and
- Results for intended beneficiaries.

It is also argued by some to mean:

- Results that can be causally attributed to an intervention.

And increasingly:

- Results to which an intervention makes a difference, contributes, or adds value.

The wording used in DAC criteria – about ‘making a difference’ – reflect this latter approach.

However, all of this debate focusses on specifiable and bounded interventions. The logic of causal analysis does not fit easily with country programmes for many of the reasons already touched on in this assessment. For example:

- How does one identify ‘final results’ when over time these are not constant: what was successful at one moment ceases to be at another?
- How does one disentangle effects that can be attributed to an intervention when many different interventions overseen by many donors are taking place simultaneously?
- How does one disentangle effects when contextual factors outside of donor control may support or undermine intervention mechanisms?

Quality criteria for development of the kind used in this chapter were designed for large projects and programmes. Many of them can be stretched and customised to country programmes as we have done in this assessment. However, ‘impact’ as commonly understood in development evaluation was not attempted by CEs in EBA’s portfolio for good reasons. That does not mean that reasonable and informed judgements are impossible. CEs have in general set out to:

- Focus their evaluations on poverty related dimensions of interventions and programmes funded by Sweden where poverty is indeed salient.
- Use macro-level data to describe poverty outcomes in the countries concerned during the period being evaluated.
- Make links as far as possible between macro data on poverty and the actual outcomes of specified interventions.
- Examine the dynamics of these programmes so as put forward explanations of observed effects.

- Aggregating different interventions (or sectors) so as to arrive at reasoned conclusion about Sweden’s contribution to poverty alleviation.

Superimposing on this, different phases over the extended period of about 20 years under consideration, the CEs have reached conclusions about the difference Sweden made to poor people. This generally relied on judgements that weigh up evidence and was informed by experience and expertise. These conclusions were of qualified success, i.e.

- Many Swedish interventions did make a difference to the poor – although not in all cases; and degrees of success varied over time for reasons that can be explained by circumstances varying from country government policies; the innate difficulties of the task e.g. after a war or in the face of political instability; and changes in aid policy and aid volumes by Sweden itself.

Establishing the ‘impacts’ of aid on country governments is if anything more challenging. CEs document a serious intention to enhance government capacities to plan and implement development programmes, reduce corruption, and encourage greater alignment with Swedish objectives around human rights and democracy. Reports show that results of these efforts have been mixed, but that Sweden has been able and willing to adapt their interventions to changing circumstances and even withdraw from programmes and countries when this was judged as necessary. CE reports contain a wealth of description and discussion of dialogue and other attempts to partner with and influence governments. This material is not systematically organised or within a comparable framework. This too is a theme that could be reviewed and synthesised across all CEs in this portfolio.

*In the DAC formulation of **impact**, making a difference to the lives of poor people, CEs addressed this extensively relying on detailed description and logical reasoning to reach conclusions. This was more successful applied to poverty reduction than to demonstrating impacts around human rights and democracy.*

Table 12 – Criterion DAC 6: Sustainability

Criteria DAC 6	Key points
Sustainability: Is the continuation and viability of interventions discussed.	<ul style="list-style-type: none">• Sustainability was not prioritised in EBA’s ToRs• It was nonetheless considered in three CEs, in some depth in only one• The criterion was applied widely: from equipment maintenance to the likely continuation of service and policy reforms• Environmental sustainability was not central in CE evaluations

Sustainability was another quality criterion that was not specified in ToRs. Sustainability was used as a core evaluation criterion in only one CE; although it was discussed and applied less systematically for the assessment of interventions in two others. Successful examples of Sweden taking action to improve sustainability were identified.

Sustainability was considered in areas that are common in development evaluation, i.e. risks of poor maintenance of facilities and equipment for example in energy or drainage. It was also considered as a criterion when assessing the likelihood of education and health progress being maintained and the sustainability of governance reforms. The reliability of judgements about the sustainability of service and policy reforms are not clear.

In explanatory terms, issues of sustainability were linked to country ownership; capacity problems and the need for capacity development; commercial viability of basic services; and the importance of creating opportunities for local participation and management by stakeholders in an intervention.

Environmental sustainability has been part of Sweden’s ‘multi-dimensional’ poverty reduction approach but environment/climate change was not central in country programmes or in their evaluation. Sustainability was more likely to be understood as sustainable economic or employment development. This probably reflects the period covered by these evaluations when there was less focus than today on climate change and environmental depredation.

Sustainability was addressed in CEs in terms of the sustainability of infrastructure, services, and policy reforms rather than in terms of environmental sustainability and climate change even though this is an element of Sweden’s ‘multi-dimensional’ poverty reduction approach.

2.3 Conclusions and implications

This exercise, applying two sets of quality criteria to the CE portfolio, helps identify interim conclusions from which it is possible to begin to draw implications. Many of these will be further elaborated by analyses in Chapter 3 which focuses less on an assessment of CE reports and more on the way quality, use and learning have been understood and operationalised throughout the entire evaluation cycle.

EBA's CE portfolio generated a wealth of information that was collected with appropriate care and which supported useful analysis and reasonable conclusions. This first-generation of EBA's CEs have also helped identify many ways in which future CEs could be strengthened. As previously suggested, the existing CEs can be viewed as a 'pilot' and a learning opportunity.

It is generally accepted in evaluation that learning rarely follows from a single evaluation. Because information is scattered across five reports set in different contexts, systematic accumulation, comparison and synthesis is needed if EBA is to realise the learning potential of this portfolio. Such syntheses could include substantive topics such as capacity development and sustainability; ways of evaluating Swedish policy objectives such as democracy and human rights; methodological and data related innovations; and strategies for engagement and dialogue with country governments. In all these cases synthesis could also help inform the design of further, more focused evaluations on these topics in future.

Several of the criteria included in the two quality frameworks used in this Chapter were either not covered or were not thoroughly covered in CE reports. This applied in particular to the criteria sustainability, efficiency and ethics. This raises a question for EBA: how far does it want to require that certain criteria are consistently covered in future CE ToRs and proposals? This would partly depend on whether these or similar schema were to be used routinely as evaluation assessment tools.

We would argue that some specifics should not be optional – like requiring evaluators to adhere to accepted ethical codes. However, requiring specific skill sets may be appropriate in some cases but not others. A move towards greater pre-specification might have quality and usefulness advantages; but may inhibit innovation as well as conflict with core values of EBA such as 'double independence'. (This is further discussed in Chapter 3.) However,

making explicit the quality criteria against which work will be assessed would be desirable as this would provide evaluation teams with clearer expectations and encourage them to take greater responsibility for their own quality assurance.⁵

Long term evaluations of country level inevitably rely on well informed judgement. The experience of evaluation teams that undertook these evaluations was therefore a great advantage. The balance between judgement and methodological rigour is a fine one. Some aspects of these evaluations would have benefitted from greater evaluation and methodological experience. The kinds of methodological expertise that was notably absent was that able to assess impacts and causal relationships in complex settings. Rigorous causal analysis is arguably not well-suited to the evaluation of multi-strand country programmes (rather than specific interventions) nor over the long term. This suggests that preferred skills are more likely to be of a qualitative kind – Theory of Change, ‘theory-based’ impact evaluations, Contribution Analysis; or of the ‘quali-quant’ kind, such as QCA or Process Tracing.⁶

Selected CE teams seemed strong on practice and more familiar with development research rather than development evaluation. There was also little evidence of familiarity with broader experience in OECD countries with country evaluations. Nonetheless some of the familiar problems of country evaluations were avoided, in particular the selection of country ‘cases’ that represent broader country programmes and linking these back to country implications.

Inadequate budgets have been recognised as weakening early CEs. Even with recent budgetary increases it is not certain that planned budgets will always be sufficient. Advance estimates of the likely costs of activities prioritised in ToRs at the time of commissioning would be one way to ensure a better balance between activities and costs. Another would be to exploit synergies across EBA’s wider study and evaluation portfolio. The CE portfolio is only a small part of EBAs work but portfolio planning that combined ‘evaluations in countries’ with ‘evaluations of countries’ could allow for economies of scale. For example, thematic priorities and methodological innovations could be inserted into CEs and testing the generalisability of CE generated hypotheses could be inserted into other EBA studies involving, for example, country-based case studies.

⁵ The topic of Quality Assurance is discussed in greater detail in the next Chapter.

⁶ It is noteworthy that not a single CE in this portfolio contained a recognisable Theory of Change, nowadays regarded as a requirement in most international development evaluations.

Country involvement in all donor driven evaluations including CEs is considered good practice. This would normally include prior consultation, taking account of country priorities in evaluation design and feeding back evaluation results. This seems to have happened to a very limited extent in the CE portfolio. In some cases where there are differences in priorities between Sweden and partner countries, a high degree of donor/country coordination might be difficult. (As noted above this topic was also prioritised by Statskontoret in its 2018 report.)

Whilst prioritising Swedish priorities is understandable, all aid occurs in settings where other bilateral and multilateral donors also operate. A consistent weakness of the CE portfolio is the failure to analyse interactions and synergies between Sweden and other donors. This limited the ability of CEs to reliably evaluate the contribution and value-added of Swedish aid. Interaction with other donors is also important in order to better understand aid governance which involves all donors and country governments and is an area where innovation will be required with Agenda 2030. This was a notable gap in coverage in this CE portfolio.

Finally, the application of two ‘quality schema’ to EBA’s CE portfolio has highlighted a number of operational challenges. First this task has highlighted the importance of streamlining elaborate assessment schemes to make them manageable. Including DAC as well as EBA’s recent assessment policy was also in part recognition that in their early stages all such frameworks undergo a period of revision and refinement.

Second, all quality frameworks are what they say they are – frameworks rather than blueprints. They require time and the application of judgement. Quite possibly different ‘assessors’ could reach different conclusions or at least emphasise them differently using the same quality framework. Familiarity using assessment schema is one way to achieve a degree of consistency. Many development agencies apply assessment schemes routinely. Application can range from self-assessment by evaluators, assessment by in-house evaluation experts and establishing a panel of external assessors who become familiar with agency priorities and also become efficient through practice. Assessing the quality of CEs, itself offers opportunities for learning and improvement.

3 Quality, Use and Learning in Practice

3.1 Introduction

The previous chapter focussed on the quality of CE reports using existing quality frameworks and criteria. This chapter broadens out to examine the way quality, use and learning were practiced by stakeholders in the CE portfolio. The word ‘practice’ is used to include the dynamics that surrounded CEs, i.e. what was done; and the way the key concepts of quality, use, and learning were defined and understood.

This chapter discusses ‘quality’, ‘use’, and ‘learning’. These terms are socially constructed – they take on meanings as a result of negotiation between stakeholders and communities of practice. The results of such discussions and negotiations vary across settings. This is because these terms are also contextualised: shaped by the demands of particular activities and the needs of specific stakeholders and organisations. Given the way those we interviewed understood quality it was also often difficult to neatly distinguish between ‘quality’, ‘use’, ‘learning’, and ‘relevance’ as these terms were often used interchangeably. Despite this we have tried to address each of these core themes in turn.

Whereas Chapter 2 applied evaluation policies and ‘standards’ to CE reports; this chapter relies mainly on interviews with individuals who were involved in CEs – as commissioners, team leaders, members of Reference Groups, and as expected users in MFA and Sida. These interviews were backed up by documentary sources where possible. However, it is worth noting that this assessment encountered problems similar to those encountered by CEs themselves. Many key individuals had moved on; others could not remember much about what for them were historic and sometimes fleeting encounters; and those now occupying key roles had few records to rely on. Indeed, the lack of documentation and minutes about key CE decisions was itself striking.

In addition to interviews and documentary sources, this chapter also draws on available research into areas such as quality assurance, evaluation use, learning pedagogics and knowledge transfer/translation.⁷ However this is not an academic text: what we understand to be key concepts are used but citations from literature are intentionally minimal.

⁷ See for example: Minkman, van Buuren and Bekkers (2018) Stone (2012).

3.2 Evaluation ‘Quality’

Defining quality is never easy.⁸ This is partly because in everyday speech as well as in dictionaries, quality is an integral ‘attribute’ or ‘characteristic’ difficult to separate out. The practice of quality is easier to define if we look to see how quality is operationalised. Whether in industry and commerce; professional practice; science and research; or in evaluation quality is usually operationalised in terms of three building blocks:

1. Quality standards that specify what is fit for purpose and reliable; and ideally differentiates between the excellent, adequate, and unacceptable.
2. A consensus process that agrees what these quality standards are and gives confidence that standards travel between interested parties such as buyers and sellers or users and producers.
3. A quality improvement process that aims to increase quality – or at least not allow standards to fall.

There was little evidence of these building blocks in EBA when CEs were being implemented. EBA’s Quality Policy had not been promulgated at the time. Even now EBA is at the very early stage of applying its recently adopted quality standards and this assessment has not been able to identify any explicit quality improvement roles or responsibilities. There has been a basic ‘quality control’ function in EBA, vested in EBA Committee itself. It is targeted at reports rather than the overall evaluation process.

A broader quality assurance approach would for example oversee the entire evaluation cycle – seeking out the preconditions or threats to quality prior to the report stage. These might include for example, the suitability of ToRs, appropriateness of budgets, capabilities and skills of retained staff, commitment to ethical codes, appropriate methodological choices; liaison with stakeholders and partner countries etc. As important a QA system would identify roles, responsibilities, and moments in time when quality scrutiny and ‘steering’ might be expected. For CEs it was only when a final report was submitted that there were formal opportunities for RG and EBA members to comment on report quality. A quality assurance system would require earlier and continuous interactions starting from expectations set at the Commissioning/ToR stage.

⁸ See Dahler-Larsen (2019).

Unsurprisingly, this assessment exercise was not able to identify consensus about what constituted quality either within EBA. Consequently, there was also no consensus about quality between EBA and stakeholders such as Sida and MFA. Those interviewed as part of this assessment seemed to understand quality mainly in ways consistent with what they thought evaluation was supposed to do. For example:

- Those who identify evaluation with judging success, goal attainment and accountability, regard evaluations to be of high quality when they support such judgements. This would be consistent with a ‘results’ orientation to evaluation and public management.
- Those who emphasise an analytical, explanatory, and methodological purpose of evaluation understand evaluation quality in terms of favoured methods, analytic sophistication or even quantification.
- Those committed to a ‘utilisation’ understanding of evaluation’s purpose often merge quality with use and utilisation – if it’s used or usable to us, it must be good.⁹
- Those who emphasise learning purposes of evaluation tend to associate good quality with the knowledge that evaluations generate hence focusing on innovative report content and the processes surrounding evaluation commissioning, production, and dissemination.
- Those who take a normative position on evaluation (emphasising the ‘value’ element) highlight both the ethics of those conducting an evaluation and adherence to Paris Declaration principles.

and finally

- Those who take an administrative/procedural view of evaluation often judge quality in terms of whether a report contains a clear executive summary; ends with conclusions and recommendations; and does not exceed a fixed number of pages.

All the above expressions of quality came up in our interviews and unsurprisingly there was considerable variability in the way the quality of any one CE was assessed by those interviewed. Often respondents were talking at cross-purposes. For example:

⁹ Unfortunately, we know that ‘bad’ evaluations may be used as often as ‘good’ evaluations.

- The same CE might be assessed as ‘interesting’ by one stakeholder because it ‘made them think in new ways’ and as ‘very poor quality’ by another because it failed to answer questions they were interested in.
- CE report might be judged as of ‘sufficient’ quality because it included a ‘Conclusions’ section or of ‘poor quality’ because it did not include a separate ‘Recommendations’ section.

As already noted, there was a lack of systematic quality assessments of CEs other than at the endpoint when draft final reports were prepared. For example:

- Stakeholder feedback was not consistent or systematic.
- Assessments could come from a geographical desk in Sida; from a senior official in MFA; or from an embassy based Sida official.
- Only three out of five CEs received something resembling a formal Management Response and there was no obligation that such responses be provided.

Quality judgements made coincided with the most common ‘quality’ concerns in many evaluations:

- The appropriateness of methodology – although such criticisms were not always soundly based.
- The strength of evidence supporting conclusions.
- Length of reports, especially important for busy officials: ‘we don’t have time to read 200 pages’; ‘we need an executive summary’.

One interesting criticism concerned the appropriateness of data collected, for example, it was suggested that:

- There should have been more in-country fieldwork and inputs from country projects; and
- CEs should have evaluated country programmes.

More country fieldwork was also a preference of Team Leaders interviewed if budgets had allowed. On the other hand, the extended time perspectives of CEs also justified greater dependence on documentary sources. There was some expectation from Sida staff in embassies that CEs should evaluate Sweden’s country portfolio. CEs would undoubtedly have been strengthened if they had been able to draw on a library of past country programme evaluations; and CEs did for the most part take

previous evaluations into account. However, as we understand it, country evaluations are not part of Sida's structured strategy planning process. To that extent this criticism of CEs also highlights a systemic evaluation anomaly: infrequent country evaluations in the overall Swedish aid system.

Finally, it was striking that although quality assessments of CEs by stakeholders were often critical this contrasted with a generally positive perception of EBA itself. EBA's work was often regarded as of good quality even if interviewees had reservations about CEs. This suggested that the authority and status of EBA of itself carries a 'quality stamp'.¹⁰

Quality assessments by stakeholders are unlikely ever to be unambiguous and definitive. Agreed and bounded definitions of quality, as in EBA's current Quality Policy and in DAC Evaluation Criteria would have been helpful if they had been in place between 2016–2019. However given the positioning of EBA in the Swedish aid system, any set of definitions would need to be supported through a deliberative, consensus building process that included not only EBA and its authors, but also the other main aid actors – in government, Parliament, MFA and Sida as well as wider publics.

3.3 Evaluation use

Producing reports that no one reads or takes seriously has been a long-held concern of both commissioners and producers of applied research and evaluation. The literature on evaluation use, usability, implementation and influence is now extensive. This literature and the way 'use' enters into evaluation practice is also diverse. A cottage industry has grown up over the last half century dedicated to defining evaluation use.¹¹ The main 'types' of use commonly identified in these literatures include:

- *Enlightenment* or *conceptual* use that encourages new thinking and innovation in policy or practice usually over the longer term.
- *Instrumental* use that sets out to answer questions and provide information to support action and decision in the shorter term.

¹⁰ This also suggests that the boundaries between CEs, other evaluations and other EBA studies was not always clear in stakeholders' minds.

¹¹ For fuller background see in bibliography: Alkin and King (2017 & 2019); Herbert (2014); Hoydal (2019); Kirkhart (2000); Mark and Henry (2004); Patton (2020); Saunders (2012); Weiss (1977 & 1979).

- *Symbolic* use that serves legitimisation or political purposes – often seen as evaluation ‘misuse’.
- *Process* use where engagement of stakeholders with evaluation supports individual and organisational learning and development.

These ‘types’ have implications at every stage of the evaluation cycle – from when evaluations are commissioned through to ‘users’ responses to recommendations. However, there is still a tendency in evaluation practice to emphasise quality at the report stage.

The above categorisations originate from within the evaluation community rather than from administrative, political-science or practice research. Unsurprisingly therefore they place evaluation and the evaluation report centre stage. They rarely start from the perspective of how policy and practice actors utilise knowledge or how institutions and organisations learn. However, this changed to a degree with the appearance of the ‘influence’ framework:

- Evaluation influence acknowledges first that evaluation is only one source of knowledge and information that policy makers, managers and practitioners rely on; and second, places evaluations within an organisational or systemic setting.

This framework begins to pay attention to the world of the ‘user’, although evaluation use/influence frameworks do not go far enough in taking account of who potential users are and the knowledge acquisition and learning strategies of policy actors such as EBA’s main stakeholders.

Evaluation ‘use’ is also associated with different visions of evaluation and its purpose. For example, possible uses of evaluation include:

- Supporting ‘democratic accountability’ by enabling better informed public scrutiny e.g. within civil society – including parliament, the academy, and the 3rd sector.
- Strengthening public-management and accountability by policy actors within the kinds of results frameworks favoured by 1990s public management ideologies (New Public Management).
- Providing timely information to steer and further develop existing programmes and policies or in a slightly tighter form answer pressing questions that stakeholders want answers to – closest to what has often been called ‘instrumental’ use of evaluation.

- Generally inform an understanding of policy choice and explain drivers and barriers to change by advancing new concepts and theories – close to Carol Weiss’s notion of ‘enlightenment’ or ‘conceptual’ use.
- Providing ‘ammunition’ for the protagonists in policy and implementation conflicts to reinforce their arguments or justify cuts in programmes or budgets– similar to ‘symbolic’ and ‘political’ use.
- Offering reassurance – the appearance of transparency – that aid policy is knowledge-based even though this might only be ‘symbolic’ i.e. there for show.

Again, all of these understandings were present among those interviewed for this assessment.

- The dominant perspective of the usefulness or otherwise of CEs among those with operational responsibilities in MFA and Sida, was the extent that CEs contributed directly to country strategy-making. It is understandable that busy front-line staff would prefer CEs which helped them with their work, i.e. facilitated ‘instrumental’ use.
- For many stakeholders the long-term 20-year view of CEs made use less likely. It was accepted that EBA had every right to commission CEs that did *not* expect to be useful in the short-term – ‘but then it’s of little use to us’.
- Senior staff interviewed were more open to the general CE insights or to ‘conceptual’ use: CEs ‘made you think’, ‘the seminar I attended was very interesting’. However even when direct use was anticipated – as was the case with the Cambodian Decentralisation study and Bosnian Herzegovina CEs it was difficult to identify what use had actually occurred. Perceptions of usefulness is not a guarantor of actual use.
- Actual use is also a matter of policy priorities and politics. The Cambodian Democracy CE was clearly seen as relevant and useful by Sida and could have been used to strengthen future strategic plans. However, as the government had already decided to close-down the Cambodian country programme the CE did not fulfil that potential.
- The organisational memory of MFA and Sida can be short-term. Some recently appointed staff responsible for Tanzania in 2020 were unaware that a Tanzanian CE had taken place in 2016. The staff responsible for promoting democracy in development cooperation in MFA had no knowledge of the Cambodian Democratisation CE – although they were involved in more recent EBA (not CE) studies.

Different explanations can be advanced for the above:

- In order for CEs to be instrumentally useful the timing and choice of country would need to be aligned with strategy cycles, but the selection of country and timing of CE reporting was not consistently aligned with strategy cycles. Country choice was usually left to CE authors choosing from a list of possible countries; and timing of upcoming strategy did not appear to be a key criterion for EBA when drafting ToRs.
- The principle of ‘double independence’ in which EBA’s priorities are seen as independent of government; and authors’ are independent of EBA in their choice of country from within a list of possibilities; and choice of theme (poverty reduction; democracy; human rights; decentralisation etc.) reduced the alignment that would have been needed for greater ‘instrumental’ use.
- Opportunities to steer CEs once commissioned in directions that would maximise policy usefulness would need management and Reference Group interventions. However, ‘policy steering’, was not seen as a core responsibility of Reference Groups members. Even now RG Guidance emphasises that RGs are expected to enable and support CEs whilst respecting the independence of CE authors. As already noted, RGs do not have a quality assurance brief.
- At the time that the CE portfolio were underway active participation of Sida and MFA representatives in RGs was not mandatory and did not always occur. However, it is noteworthy that when Sida in particular was represented on RGs these representatives did press for greater policy relevance and better alignment in terms of focus and timing.
- Liaison and coordination arrangements with MFA and Sida had not been as well-developed during the CE implementation period as they now are. Engaging potential users from the outset is likely to encourage all varieties of evaluation use. This would for example have to extend beyond geographical departments and desks to include those responsible for priority areas such as human rights and democracy; poverty reduction and climate change. And for some use possibilities other civil society networks would also be implicated.

However evaluation use is understood, evaluation outputs have to be communicated to be used. EBA's communication, liaison and coordination strategies with stakeholders were not well developed in 2015/2016. Nonetheless some aspects of later communications policies were already evident in the earliest CEs. For example, EBA invested in country feedback meetings in Tanzania and achieved extensive press coverage of the same CE report in the Swedish press. EBA's present communication strategy was beginning to take shape in the second stage of CE implementation.

In its current form, EBA's strategy to promote 'use' is concentrated at two moments. First when planning evaluations by identifying stakeholder in Sida and MFA with whom we were told liaison is maintained throughout the evaluation. Second, when delivering evaluation reports, the emphasis is on dissemination – to embassies, MFA and Sida counterparts in the first instance; and then when reports are published with seminars and press briefings to wider publics as well as mention in EBA's Annual Reports. The emphasis appears to be to 'get the message out'. In these terms EBA's Comms Strategy has become increasingly effective both in the latter stages of CE implementation and since.

However different communication strategies are suited to different potential uses of evaluation. If the aim is that knowledge and insights continue to be available to inform future thought and action, a more extended and interactive communication strategy would be needed.

Communications has been discussed here in the context of evaluation use. Communication as a process is also a key part of what is meant by 'learning' and is therefore referred to further in these terms in the next section.

3.4 Learning through CEs

In any discussion of learning in evaluation, it is important to clarify who is expected to learn, about what, through what modes and in what settings. Although the bulk of this section is concerned with learning by actual and potential learners both from among EBA's stakeholders and the wider policy-shaping community, the section begins by considering those who should be learning the most from CEs, the teams that undertake evaluations.

3.4.1 How evaluation teams learn

Often in evaluation discourse it is assumed that learning is what others do! It is stakeholders and target groups that should be informed by evaluations. It is also true and was evident from this assessment that learning through CE evaluation was critical for the evaluation teams responsible for the CEs.

Evaluation teams if they are to succeed in their work need to undergo their own learning process. This begins when accessing data and documentation, making contact with key informants, establishing and maintaining relationships, testing out new ideas as they begin to take shape, and refining and contextualising conclusions. EBA management are key gatekeepers here, especially by making contacts in MFA and Sida; and via Stockholm headquarters with embassy contacts in-country. Although EBA strengthened these liaison arrangements during CE implementation and since, interviewees agreed that these arrangements were not optimal at the time.

Team leaders reported many problems of access and communication which had implications for their own effectiveness and learning:

- CE's required access to country plans, project documentation, statistics, and where available relevant evaluations. Access to documentary sources via Sida and MFA was also not always easy although this improved in later CEs. In one early CE, after repeated unsuccessful requests in Stockholm, the CE team had to access statistics on Swedish aid via OECD.
- Team leaders described difficulties accessing individual gatekeepers and informants in MFA and Sida. It was suggested that if stakeholders decided that an evaluation was not useful or timely, they were less likely to prioritise CE requirements.
- Difficulties were greatest in embassies, for whom sometimes the appearance of a CE was a surprise: 'the evaluation was like a spaceship coming from cyberspace'. CEs that conducted fieldwork in-country often relied more on their personal contacts and networks than on Embassy connections.

Learning during the course of an evaluation is closely related to what is called ‘process’ use in evaluation parlance. It occurs in stages, i.e. when evaluators are:

- scoping, planning and negotiating access
- collecting information
- cross-checking and validating analysis and findings
- contextualising conclusions and recommendations
- disseminating reports and customising dissemination for different audiences

Contact, dialogue, and exchange with stakeholders for CEs was constrained by budgets and reinforced reliance on documentary rather than in-country fieldwork. This limited the extent of dialogue between CE teams and MFA, Sida and Embassies. Reactions to CE reports and the content of these reports would have been improved if such dialogues had been established early and sustained. For example, validation meetings with Embassies did not routinely take place although dissemination meetings in-country were more common. This in our view weakened mutual learning by CE teams and stakeholders and ultimately the relevance and usefulness of CEs. There seemed to be no expectation communicated by EBA that feedback would be provided to partner governments or country stakeholders.

The next sections focus more explicitly on pedagogics: the processes and modes of learning necessary for knowledge acquisition. In addition to what would conventionally be understood as ‘learning theory’ and pedagogics, the section is also informed by policy ideas about policy and organisational learning and knowledge acquisition.¹²

3.4.2 Learning as ‘transmission’, ‘co-production’ and ‘translation’

The default assumption in evaluation is that ‘learning’ is the transfer of knowledge from those who know to those who need to know. In pedagogic terms the assumed mode is *transmissive* – we must identify the putative learner and give them information so that they can change their behaviour. Implicitly transmissive learning in evaluation resembles what

¹² See for example: Ciborra (2002); Nonaka (1994).

psychologists call ‘one shot learning’: a single output (launch of a report) being judged sufficient for knowledge transfer. Although EBA also refers to ‘conceptual’ and ‘process’ uses of evaluation which imply a more differentiated notion of learning, the transmissive mode also seems to predominate in EBA.

This default mode may be suited to short-term or immediate use of evaluation by identifiable individuals with pre-identified knowledge requirements e.g. in order to draw up better country strategies. Even in these circumstances a transmissive mode of learning is likely to be most effective when what is being transmitted are data or facts rather than, say, principles or concepts. For example, lessons about country programme implementation or the ways democratic practices can be integrated into economic development do not sit easily within a ‘one-shot’ learning approach. As argued above, new or challenging ideas require interrogation and dialogue in order to be internalised and contextualised.

3.4.3 Alternative ‘learning scenarios’

This assessment has identified a range of circumstances within which evaluations might be expected to contribute to learning. We were not however able to find much ‘evidence of learning’ other than assertions that a particular CE was ‘interesting’ and ‘we’ll come back to it’. However, we were able to clarify some of the main taken-for-granted assumptions about learning, and a broader set of potentially usable pedagogic practices than transmissive information transfer alone.

These are described below as ‘learning scenarios’. These scenarios are based on a number of different but well recognised pedagogic practices. They include but are not confined to ‘transmissive’ learning. Other modes include ‘collaborative learning’ and the ‘co-production’ of knowledge; ‘policy translation’ across institutional settings; and ‘dialogical learning’ in and through knowledge networks.

The rationale for these scenarios prepared is heuristic. They are intended to exemplify a broader range of pedagogic modes than we in fact encountered. With regard to ‘learning’ we discovered that we were being asked to assess a potential rather than actuality. The suggested scenarios combine empirical, conceptual, and imagined material. Generating plausible ‘imaginaries’ that illustrate how learning through evaluation might be improved is a response to the lack of accessible evidence that CE related learning took place; or of the existence of explicit learning strategies in EBA at the time.

Each scenario describes categories of learners; what they might be expected to learn; the mode of learning implied; and the setting institutional or otherwise within which such learning might take place. The first scenario is close to the default ‘transmissive’ mode of pedagogy as already outlined in this section. Following scenarios are at least implied by evidence gathered as part of this assessment and discussions about quality, use, communication, and learning throughout this report.

Scenario 1: Knowledge Transfer

Following consultation there has been agreement between EBA and Sida about a topic of interest: to evaluate on a pilot basis methodological innovations when planning country strategies. Sida is interested to know more about the effectiveness and value-for-money of using ‘frontier’ methods such as Bayesian Updating; adaptive programming; and Process Tracing. A particular question is to evaluate the appropriateness in such innovations in different country and strategic contexts.

As a result of prior consultation, a set of potential users (based both in Stockholm, one regional office and in 4 Embassies) were identified before the evaluation was commissioned. Sida have agreed to help the retained CE team by providing access and background information.

The basic pedagogic mode here is transmissive – the planned CE is expected to produce usable results that Sida will take up and use in its own way. Learners are known and the institutional setting is known. However, the team undertaking this work will have to understand the enablers and barriers to learning at different administrative levels as well as in different country settings.

Scenario 2: Collaborative learning and Co-Production of knowledge¹³

MFA has expressed an interest in scoping out a new strategic priority to reduce domestic violence. As smaller scale initiatives are already present as part of existing development cooperation strategies, the Ministry wishes to evaluate this experience. EBA has been asked to develop and test an ‘evidence framework’ that could be used at a country-level; and work with one regional office and one embassy as ‘proof of concept’. The justification for EBA’s involvement is the importance of having an independent and robust framework that has credibility but still leave policy options open.

¹³ See Armitage et al (2011); Swedlow (2010).

As in Scenario 1 there is a prior commitment to using and learning (i.e. there are known users), although the content of what might be learned is less pre-determined and may vary across country settings and regions. What is distinctive here is that the evaluation team contracted by EBA will need to work closely with officials and experts. Knowledge to be usable will have to be co-produced. A dissemination strategy will be required to support cross-embassy learning. Although EBA will not be responsible for dissemination among Embassies. EBA will need to work with the retained consultants to ensure that in addition to technical report, targeted communications or mini-reports and tools are also prepared.

This scenario also assumes known users/learners but relies on collaborative learning and the co-production of knowledge as a primary pedagogic mode. The scenario also addresses intermediate rather than short-term learning timescales and does so in a differentiated but identified institutional setting. Although this scenario anticipates ‘use’ and ‘learning’ the outputs are intermediate rather than immediately used and ‘instrumental’.

Scenario 3: Policy translation and institutionalisation

Following a recent mainly statistical study of international experience of democracy promotion, EBA has decided that further evaluation and research is needed into how to reconcile Sweden’s Democracy and Human Rights strategy in partner countries that do not themselves support these priorities. The aim is to supplement recently completed statistical impact studies, with a case-study evaluation of implementation arrangements associated with successful outcomes. These case-studies will focus on the most promising initiatives funded by Sweden, other DAC countries and multilaterals including UNDP and the EU.

Following consultation, this evaluation has been judged by the government as being politically sensitive. Nonetheless EBA has decided that this should be prioritised in their workplan to contribute an independent perspective to future policy development. Whilst there is no immediate identifiable ‘user’ or ‘learner’ MFA has agreed to be part of the membership of a ‘quality assurance group’ being set up by EBA and it is expected that the Division for Democracy and Development Cooperation will also provide inputs.

Research into policy learning in institutional settings highlights the importance of ‘policy translation’ rather than earlier theories of ‘policy borrowing’; and the mutually ‘constitutive’ effects of knowledge and

policy.¹⁴ (Previous dominant theories of ‘policy borrowing’ strongly resembled a ‘transfer’ approach to learning.) Those undertaking this study will be expected to use a policy translation framework to explore the potential for learning across different donor/Development Cooperation policy and institutional settings. It is hoped that resulting policy translation protocols and methods, can be useful in future to promote policy learning across different sectors.

In this scenario there are no signed-up learners but a potential learning community exists and this evaluation has the potential in the medium term to mobilise potential learners across the Swedish aid system. In support of such a scenario, a more extended dissemination exercise would be needed. For example in addition to a single report there will be a number of targeted mini-reports and technical briefs; and in addition to a single launch event there will be a commitment to present outputs at conferences and practitioner and professional development programmes – including those organised for development practitioners.

Scenario 4: Knowledge networks

A recent assessment of EBA’s previous CE portfolio identified a number of ‘themes’ that merited further evaluation and synthesis. These themes include strategies for engagement with country governments, capacity building, integrating sustainability across development projects; and coordination with other donors including multilaterals. Whilst some added value will follow from thematic syntheses across CEs and other EBA studies and evaluations this will not be sufficient for learning purposes. Many of these themes intersect with academic areas of study and research; and with practice interests among agencies that have as part of their mission implementation and change agent roles. EBA therefore decided that the most effective way to maximise the potential learning yield from its investment, was to initiate a number of ‘knowledge networks’ in collaboration with university and other public agency sponsors.

These networks will engage with relevant policy-shaping communities: researchers, knowledge brokers, practitioners as well as policy makers. They can be regarded as ‘sites’ for dialogue. Networks such as these rely on shared learning, collaborative learning, and the co-production of knowledge. It is hoped that they will also contribute to ‘knowledge translation’ across communities of practice.

¹⁴ See especially Christina Boswell, Diane Stone and Peter Dahler-Larsen.

Networks can take different forms perhaps starting with monthly or bi-monthly Webinar study groups reinforced by a larger annual face-to-face day-long workshop. If designed appropriately – for example around an inclusive meta-theme, ‘implementation and uncertainty’ was suggested – knowledge networks could become a useful infrastructure for EBA more generally, providing a dissemination and engagement vehicle for a variety of EBA outputs.

This scenario does not have pre-identified ‘learners’: the aim is to identify and recruit networks of collaborative learners. The mode of learning relies on dialogue and co-production. EBA’s role is one of enabling dialogue and knowledge integration. This scenario is not institutionally located but can be seen more as a ‘boundary-spanning’ vehicle for knowledge sharing.

3.5 Conclusions and Implications

The words ‘quality’, ‘use’, and ‘learning’ like most terms in evaluation can be interpreted in different ways; and sometimes words are used interchangeably. This instability of language was evident at the time CEs were being implemented.

There are, for example, many meanings attributed to the term quality in evaluation, sometimes overlapping with notions of usefulness, sometimes dependent on methodological rigour, sometimes on stakeholder engagement; and sometimes depending on whether an evaluation’s findings or recommendations do or do not agree with an official’s own view or current government policy. There was no EBA Quality Policy until 2020. Having a policy that is clear, well communicated, steers evaluator behaviour can help shape common expectations provided the policy is implemented. This depends on some kind of comprehensive ‘quality assurance’ process, present at every stage in the evaluation process from commissioning onwards; and supported by roles responsibilities and procedures. Given the recency of EBA’s Quality Policy it is perhaps unsurprising that such a broader QA system was not in place when CEs were being implemented.

The Specification of this assessment exercise asked apparently straightforward questions such as: ‘Have reports contributed to the development of Swedish development cooperation?’ and ‘who is learning from the country evaluations?’. Answering such questions is not straightforward. For four out of the five CEs we can find positive judgements about use and learning in memos from ministry officials;

feedback from stakeholder representatives on some Reference Groups; and interviews conducted for this assessment. We also encountered negative judgments about the same CEs, sometimes from the same officials or from their colleagues.

An individual saying a report is useful and offers lessons does not mean it is used; and within an admittedly resource-limited exercise we were unable to find evidence of use or learning even when this was anticipated in 2016 or 2018. If EBA was an ‘internal’ evaluation unit within an agency or ministry, one might advocate the advantages of formal Management Responses.¹⁵ Whilst some kind of follow-up mechanisms appropriate to EBA’s independent status *would* be beneficial, this would not overcome the problems that follow from the lack of a shared evaluation lexicon.

Moves to further develop and institutionalise EBA’s Quality Policy offers opportunities to implement a shared lexicon for evaluation quality and overlapping terms such as use and learning. However, in a multi-agency setting a set of standards, meanings and judgments cannot be confined to only one of the partners. Ideally one would aim for a consensus-based lexicon arrived at collaboratively. At the very least it would be important for EBA to communicate both to its authors (evaluation contractors) and to its stakeholders its understandings of what learning, quality and use mean. This would make it more likely that all parties converged in the way they formulated and used evaluation outputs.

We have argued that the default notion of use in EBA is ‘instrumental’ i.e. that CE findings should be taken on board to modify decisions and even policies in the short-term. This has not happened and is not likely to. It would be unusual in most policy evaluation settings. However, at the time of these CEs and to a large extent also now, EBA’s arrangements and procedures convey mixed messages. CEs were not consistently aligned with the timing of upcoming preparation of country strategic plans. And even when alignment appeared possible the ‘double independence’ principle that offered authors a list of alternative countries could derail potential alignments. At the same time EBA and stakeholder informants suggested that CEs might indeed have achieved ‘instrumental’ use and led to policy ‘learning’ if only coordination and liaison had been better when CEs were launched. This is partly a question of expectation management.

¹⁵ A complementary approach would be for EBA, MFA and Sida to jointly undertake periodic ‘follow-up’ audits to clarify to what extent and in what ways CEs and associated studies have influenced development policies and practice.

When senior officials in MFA fully accepted that a CE was not necessarily directed to short-term priorities, these expectations were not always shared by front-line officials. The latter hopeful of support sometimes formed their own expectations that the CE *would* be useful to them. When this proved not to be the case front-liners could easily become resentful and critical. This may explain why CE teams reported receiving only very limited assistance from in-country officials. Distinguishing clearly between those CEs that are expected to support instrumental use and those that are to have longer term or more diffuse use and learning purposes would be helpful. It is also important that expectations about evaluation support, advice and information input are communicated consistently to geographical departments and Embassy staff.

It is likely that the current extent of use and learning of CEs and other outputs is under-recorded. We know that EBA's seminars have been well-attended; those attending from many walks of life have found them stimulating. Sometimes there has been press comment on reports; and EBA's Annual Report is widely read. From a public accountability and scrutiny perspective CE reports have undoubtedly added value. Senior officials in particular have found engagement with EBA including its CEs as contributing to their learning. Some have said referring to CE reports, that they 'will come back to them in the future'. If we move away from a short-term 'instrumental' definition of report use, designed to answer pressing questions and backed up by transmissive communication and dissemination strategies, then there probably has been more use and learning than would at first appear.

At present, the predominance in EBA of 'transmissive' ideas of learning may crowd-out consideration of other more collaborative, networked and translational understandings of what learning can mean in a multi-agency policy environment such as EBA's. Because learning is not only about transmitting new facts, more diverse pedagogic and policy translation modes of learning (and communication) need to be considered. Even though these more diverse modes are not currently acknowledged, they are probably immanent, having the potential to fit better with CEs and probably other strands of EBA's work. It is for this reason that this chapter also elaborated a number of additional 'learning scenarios' to illustrate what a more diverse approach to learning – and use – might look like.

4 Looking Forward

4.1 Introduction

It has been emphasised to us that this assessment should lead to ‘concrete recommendations’ about how EBA can improve its work. This Chapter is a little less directive than that in terms of ‘recommendations’, for two reasons. First it is never entirely obvious that external assessors or evaluators are the best people to recommend specific actions that have to work in a setting that as outsiders, they may not fully understand; and second recommendations (in contrast to conclusions), usually benefit from joint discussion and feedback with those responsible for implementation before being committed to text in any detail.

Those who read Chapters 2 and 3 will already have encountered a swathe of suggested actions – the menu is potentially extensive. Rather than repeat or further concretise these suggestions here, we identify an outline ‘agenda for CE strengthening’ focussed around broad areas in which in our view, something needs to be done.

In this agenda, we go so far as to indicate the kinds of ‘somethings’ that might work, but they are framed with a greater degree of flexibility and openness than standard recommendations. This is partly in recognition that we are not fully cognizant of what may already be happening or planned in EBA. We are open to further expanding on aspects of this agenda if this is requested after discussion. However, we can be definitive about one thing. In our view a comprehensive and sustained ‘agenda for CE strengthening’ is needed if future CEs are to realise their potential to support ‘well-founded decisions’ in this important aspect of Swedish development cooperation.

Finally, the agenda as outlined below is indicative. Others having read this report may prioritise different follow-up actions. The options outlined below indicate how the challenges identified in this assessment could be addressed. It is also possible that there are other ways to achieve the same ends.

4.2 An outline agenda for CE strengthening

4.2.1 From quality policy to quality practice

EBA's recently published Quality Policy will probably need further development but this would best be in tandem with developing quality as practice. Quality is used here in an inclusive way as in the current EBA policy, encompassing principles, criteria, an assessment process as well as roles and responsibilities.

Practice would encompass a Quality Assurance (QA) system that focused on the entire evaluation cycle – from ToRs, through to selection, commissioning, process management, liaison with stakeholders, problem solving etc. This would ensure that the pre-requisites of quality were in place and managed, including for example, clear purposes, partners and 'users' aligned with timing and country selection; the suitability of skills and budgets to match planned activity; agreed access to data and field sites; gate-keeping and liaison in-country etc. It would include 'Quality Control' at report stage but would not be confined to quality control.

Responsibilities for a QA system would need to be integrated and distributed. It would have implications for evaluation teams, EBA's Secretariat, Reference Group members and EBA itself. Such a system would also need to be supported by EBA's stakeholders as they will be implicated in implementation.

4.2.2 Priority setting, evaluation and portfolio planning

A future set of CEs need to be planned at portfolio as well as individual evaluation level. Some CEs may address long-term strategic goals; others may have short-term purposes. Some may be free-standing and others implemented in partnerships. This kind of planning would need to align with government priorities and stakeholder plans.

What is envisaged here would be separate from but set the parameters for operational tasks such as evaluation management and Quality Assurance. One requirement would be to set and communicate clear expectations. If a CE is intended to be 'instrumental' and useful in the short term this needs to be understood and followed through in terms of timing, selection of evaluation team, country choice and topic to be evaluated; and

supported by appropriate management and communication strategies. This would be quite different in profile from a ‘conceptual’ evaluation, for example focussing on engagement strategies with country governments or aid governance or Swedish aid effectiveness in a multi-donor context.

Portfolio planning may also have implications for other parts of EBA’s overall study portfolio as evaluations with in-country elements (e.g. case-studies) can contribute to insights about particular country programmes; just as what might be primarily a CE can also support methodological or thematic insights.

4.2.3 Knowledge accumulation and follow-up

Learning rarely follows from a single evaluation. Knowledge accumulates across cases and contexts. Existing CEs contain many tantalising insights about topics such as capacity development and sustainability; ways of evaluating Swedish policy objectives such as democracy and human rights; methodological and data related innovations; and strategies for engagement and dialogue with country governments. To maximise the yield of existing CEs some kind of synthesis and integration of findings would be needed. Such an exercise may first allow for some firmer lessons to be drawn from existing CEs; and by narrowing down gaps in understanding it may also point the way for future CEs and thematic studies.

This assessment has suggested that it is unclear whether knowledge described as ‘useful’ is ever used. Given the importance of improving CE’s lesson learning potential, it would be helpful to conduct a follow-up assessment, in cooperation with MFA and Sida, a couple of years after the completion of an individual CE. This would aim to trace how the outputs of past CEs (and possibly a subset of other EBA studies) have or have not had influence. Identifying the mechanisms or arrangements that encourage or discourage uptake would be especially valuable.

4.2.4 Communication strategies for learning networks

This assessment has suggested that EBA – and MFA – have a tendency to default to ‘instrumental’ uses of evaluation e.g. making inputs to prepare imminent country strategy plans or support pressing problem-solving. This default mode of use, is associated with ‘transmissive’ learning approaches, communicating to targeted learners and users. EBA’s communication strategies are well-developed to support this default scenario.

The assessment has suggested that other forms of ‘use’ and ‘learning’ are appropriate requiring other learning scenarios, e.g. when knowledge may have to be co-produced with development practitioners rather than by evaluations alone; and when networks are the most appropriate setting within which learning can occur. Networks are especially important when knowledge and understanding require communication over an extended period of time; and when potential learners are dispersed rather than located in a single institution.¹⁶ It is also likely that in these settings individuals will not immediately identify themselves as ‘learners’, partly because it takes time to become an active learner, able to translate or customise knowledge to fit into different settings. Supporting such networks requires a more dialogical and sustained form of communication than in a transmissive/instrumental learning scenario.

Evolving EBA’s communication strategies in this way would extend the communications repertoire to be more able to support the dissemination of innovative, conceptual, and challenging ideas over the longer term.

¹⁶ For example, networks might draw together individuals from multiple institutions, from civil society including NGOs, as well as from universities and political parties.

References

Documents Reviewed

- AidData (no date). Proposed Impact Evaluation of a SIDA-Sponsored Public Infrastructure and Local Governance Programme in Cambodia.
- Andersen, H. et al. (no date). Proposal: Evaluation of Swedish long-term development cooperation with Cambodia.
- DEval (2019). Country Portfolio Reviews. A tool for strategic portfolio analysis in German development cooperation 2019.
- DEVFIN Advisers (2017). Proposal to The Expert Group for Aid Studies. Evaluation of Swedish long-term development cooperation at country level. September 17, 2017.
- EBA (2015). Terms of Reference for proposal: Evaluation study of Swedish Long-Term Development Cooperation at country level. 2015-04-14.
- EBA (2016). Towards an Alternative Development Management Paradigm? Report 2016:7.
- EBA (2016). Exploring Donorship – Internal Factors in Swedish Aid to Uganda. Report 2016:09.
- EBA (2016). Swedish Development Cooperation with Tanzania – Has it helped the Poor? Report 2016:10.
- EBA (2017). Invitation for proposals: Evaluation of Swedish long-term development cooperation at country level. Stockholm 2017-06-30.
- EBA (2017). Guidelines for reference groups. 2017-08-28.
- EBA (2017). The EBA Aid Review.
- EBA (2018). Who makes the decisions on Swedish aid funding? Report 2018:05.
- EBA (2018). Nation Building in a Fractured Country: An Evaluation of Swedish Cooperation in Economic Development with Bosnia and Herzegovina 1995–2018. Report 2018:10.
- EBA (2018). The EBA Aid Review 2018.
- EBA (2019). EBA Methods and Procedures.
- EBA (2019). Operational Strategy.

- EBA (2019). Supporting State-Building for Democratisation? A Study of 20 years of Swedish Democracy Aid to Cambodia. Report 2019:03.
- EBA (2019). Building on a Foundation Stone: The Long-Term Impacts of a Local Infrastructure and Governance Program in Cambodia. Report 2019:4.
- EBA (2019). The EBA Aid Review 2019. Annual report from the Expert Group for Aid Studies for the year 2018.
- EBA (2020). Policy samt riktlinjer för säkerställande av kvalitet i studier och utvärderingar. (Policy and guidelines for quality assurance of studies).
- EBA (2020). Invitation for proposals: Assessment of the EBA's country evaluations: Quality, use, learning and relevance for Swedish development cooperation.
- EBA (2020). The EBA Aid Review 2020.
- Government of Sweden (2016). Policy framework for Swedish development cooperation and humanitarian assistance. Government Communication 2016/17:60.
- Government of Sweden (2017). Guidelines for Strategies in Swedish Development Cooperation and Humanitarian Assistance.
- Government of Sweden (2017). Policy for global development in the implementation of the 2030 Agenda. Government Communication 2017/18:146.
- Kruse, S.-E. (2015). Proposal. Evaluation Study of Swedish Long-Term Development Cooperation in Uganda. June 2015.
- McGillivray, M. (2015). Proposal for the Swedish Expert Group for Aid Studies. Evaluation of Swedish Long-term Bilateral development Co-operation with a Partner Country. May 2015.
- Ministry for Foreign Affairs (2003). Country strategy for development cooperation with Cambodia. January 2002–December 2006.
- Ministry for Foreign Affairs (2006). Strategy for development cooperation with Bosnia and Herzegovina. January 2006–December 2010.
- Ministry for Foreign Affairs (2006). Regional strategy for development cooperation with Tanzania. 2006–2010.
- Ministry for Foreign Affairs (2009). Strategy for development cooperation with Uganda. March 2009–December 2013.
- Ministry for Foreign Affairs (2010). Strategy for development cooperation with Bosnia and Herzegovina. January 2011–December 2014.

- Ministry for Foreign Affairs (2012). Strategy for development cooperation with Cambodia. January 2012–December 2013.
- Ministry for Foreign Affairs (2013). Results strategy for Sweden’s international development cooperation in Tanzania. 2013–2019.
- Ministry for Foreign Affairs (2014). Results strategy for Sweden’s reform cooperation with Eastern Europe, the Western Balkans and Turkey. 2014–2020.
- Ministry for Foreign Affairs (no date). Strategy for Sweden’s development cooperation with Uganda 2018–2023.
- OECD (1999). Evaluating Country Programmes. Vienna Workshop, 1999.
- OECD (2017). Results in Development Co-operation. Provider Case Studies: Sweden.
- OECD (2019). OECD Development Co-operation Peer Reviews. Sweden 2019.
- OECD/Deval (2019). Minutes. Workshop on Country Programme Evaluations and Reviews. 28.2.-2019–1.03.2019.
- Sida (2018). Evaluation at Sida. Annual Report 2018.
- Sida (2020). Evaluation at Sida. Annual Report 2019. Sida Studies in Evaluation. 2020:1.
- Sida (2020). Evaluation of the application and effects of a Human Rights Based Approach to development: Lessons learnt from Swedish development cooperation. What works well, less well and why? Volume I: Final evaluation report. Sida Evaluation 2020:2A.
- Sida (2020). Evaluation of the application and effects of a Human Rights Based Approach to development: Lessons learnt from Swedish development cooperation. What works well, less well and why? Volume II: Final case study reports – Albania, Cambodia, Colombia, Kenya. Sida Evaluation 2020:2B.
- Sida (2020). Meta-study of Partner-led Evaluations 2019. Final Report. Sida Studies in Evaluation. 2020:2.
- Statskontoret (2018). Översyn av Expertgruppen för bistandsanalys. 2018:16.
- Tilläggsdirektiv till Expertgruppen för bistandsanalys (UD 2013:01). Kommittédirektiv 2016:71.
- Utrikesdepartementet (2008). Samarbetsstrategi för utvecklingsarbetet med Kambodja 2008–2010. 2008-05-06.

Utrikesdepartementet (2014). Resultatstrategi för Sverige internationella bistånd till Kambodja 2014–2018. 2014-08-14.

Utrikesdepartementet (2014). Resultatstrategi för Sveriges internationella bistånd i Uganda. 2014–2018.

Utrikesdepartementet (2020). Strategi för Sveriges utvecklingssamarbete med Tanzania. 2020–2024. 2020-01-30.

Selected Bibliography

Alkin, Marvin, and Jean King (2017). Definitions of Evaluation Use and Misuse, Evaluation Influence and Factors Affecting Use. *American Journal of Evaluation* 2017, Vol. 38(3) 434–450.

Acemoglu, D, and J. Robinson (2000). Why Did the West Extend the Franchise? Growth, Inequality and Democracy in Historical Perspective. *Quarterly Journal of Economics* 115(4). pp. 1167–99.

Armitage, Derek, et al (2011) Co-Management and co-production of knowledge: Learning to adapt in Canada's Arctic. *Global Environmental Change* 21 (2011) 995–1004.

Cartwright N (2007) *Hunting Causes and Using Them: Applications in Philosophy and Economics*. CUP.

Boswell, Christina and Katherine Smith (2017). Rethinking policy 'impact': four models of research policy relations. *Palgrave Communications* | DOI: 10.1057/s41599-017-0042-z.

Ciborra, Claudio (2002). *The Labyrinth of Information: Challenging the Wisdom of Systems*. Oxford University Press.

Dahler-Larsen, Peter (2019), *Quality*. Palgrave Macmillan, Cham.

Forss, Kim et al. (2021) *Long Term Perspectives in Evaluation*. Routledge London and New York.

Herbert, James (2014) Researching Evaluation Influence: A Review of the Literature. *Evaluation Review* Vol. 38(5) 388–419.

Minkman E. van Buuren M. W. & Bekkers V. J. J. M, (2018) Policy transfer routes: an evidence-based conceptual model to explain policy adoption, *Policy Studies*, 39:2, 222-250.

Nonaka I. (1994). A dynamic theory of organizational knowledge creation. *Organ. Sci.* 5 14–37.

- Syrstad Høydal, Øyunn (2019). Evaluation influence: Method and societal phenomenon. *Evaluation*, 2019, Vol 25 (3) 381–398.
- King, Jean and Marvin Alkin (2019) *The Centrality of Use: Theories of Evaluation Use and Influence and Thoughts on the First 50 Years of Use Research*.
- Kirkhart KE (2000) Reconceptualizing evaluation use: An integrated theory of influence. *New Directions for Evaluation* 2000: 5–23.
- Mark MM and Henry GT (2004) The mechanisms and outcomes of evaluation influence. *Evaluation* 10: 35–57.
- Patton, Michael Quinn (2020). *Evaluation Use Theory, Practice, and Future Research: Reflections on the Alkin and King AJE Series*. *American Journal of Evaluation* (Online).
- Saunders M (2012) The use and usability of evaluation outputs: A social practice approach. *Evaluation* 18(4): 421–36.
- Stern, Elliot (2015), *Impact Evaluation: A Guide for Commissioners and Managers*. London, Big Lottery Fund, Bond, Comic Relief and the Department for International Development.
- Stone, Diane (2012) Transfer and translation of policy, *Policy Studies*, 33:6, 483–499.
- Swedlow, Brendon (2010). *Four States of Knowledge: A Cultural Theory of the Co-production of Science, Social Order, and Scientific, Cultural, and Policy Change*. Conference Paper. available at: <http://ssrn.com/abstract=1673521>
- Weiss CH (1979) The many meanings of research utilization. *Public Administration Review* 39(5): 426–31.
- Weiss CH (1977). *Research for Policy's Sake: The Enlightenment Function of Social Research*. *Policy Analysis*, Vol. 3, No. 4, pp. 531–545.

Individuals interviewed¹⁷

Lennart Peck

Roger Karlsson

Claes Lindahl

Stein Erik Kruse

Mark McGillivray¹⁸

Henny Andersen

Sven Olander

Markus Burman

Brad Parks

Karin Snellman

Eva Areskoug

Ditte Egnell

Vera Mörner

Frida Bohman Aguouram

Lisa Curman

Torgny Svenungsson

Kim Forss

Per Trulsson

Ulrika Lång

Nina Solomin

David Holmertz

Ellinor Hellberg

Marie Bergström

Jan Pettersson

Moa Bergman

Eva Lithman

¹⁷ A further 8 individuals were contacted but either did not respond or were not available during the timescale of this assessment.

¹⁸ Written response to interview questions.

Appendix 1: EBA Comments and Planned Actions

The table below lists the authors' observations together with EBA's understanding and planned actions. Not all comments require specific actions from EBA. The concrete activity related to the actions are in terms of highlighting an adding text in relevant internal steering documents (e.g. the templates for country evaluation memos and terms of references, guidelines for chairs of reference groups).

	Observation	Reference to the report	Comments and potential change of methods and procedures
Preparatory work and planning of country evaluations	Planning and choice of country for evaluation. The consideration of strategy cycles in the choice of country to include in the evaluation has been limited.	<p>"In order for CEs to be instrumentally useful the timing and choice of country would need to be aligned with strategy cycles, [...] Country choice was usually left to CE authors choosing from a list of possible countries; and timing of upcoming strategy did not appear to be a key criterion for EBA when drafting ToRs"</p> <p>"For many stakeholders the long-term 20-year view of CEs made use less likely. It was accepted that EBA had every right to commission CEs that did not expect to be useful in the short-term – 'but then it's of little use to us."</p>	In the planning of a new country evaluation, both the perspective of the portfolio of evaluations and the purpose of the single evaluation should be considered. The EBA Secretariat has initiated a process to regularly receive updated information on strategy cycles. However, not all evaluations will have the purpose to directly feed into a strategy process. The purpose and intended use of the evaluations should be clearly described in the planning phase and be written out clearly in the memo before a new study is initiated.

Observation	Reference to the report	Comments and potential change of methods and procedures
<p>Consultations with partner countries in the planning phase have been limited.</p>	<p>“It does not appear that partner countries were consulted or involved in CE approval and planning. This may also have contributed to access and efficiency problems.”</p>	<p>Consultation with embassy personnel in the planning phase is today a practice that should continue. Consultations with partner country stakeholders are not always relevant.</p> <p>In the initial planning phase, make an internal list of stakeholders that should be contacted in the start-up phase and identify key persons for the study to liaise with.</p>
<p>Design of terms of reference. Evaluation quality criteria have not been included in terms of reference (ToR).</p>	<p>“Several of the criteria included in the two quality frameworks used in this Chapter were either not covered or were not thoroughly covered in CE reports. This applied in particular to the criteria sustainability, efficiency, and ethics. This raises a question for EBA: how far does it want to require that certain criteria are consistently covered in future CE ToRs and proposals?”</p>	<p>A reference to the EBA quality policy from 2020 should always be included in ToRs. However, all DAC evaluation criteria are not likely to always be included. EBA will continue to leave room for the evaluators to design evaluations.</p>

	Observation	Reference to the report	Comments and potential change of methods and procedures
	<p>Costs have not been estimated systematically in relation to the study questions in the ToRs. Limited budgets have also limited the possibilities for dialogue and fieldwork.</p>	<p>“Advance estimates of the likely costs of activities prioritised in ToRs at the time of commissioning would be one way to ensure a better balance between activities and costs.” “Contact, dialogue and exchange with stakeholders for CEs was constrained by budgets and reinforced reliance on documentary rather than in-country fieldwork. This limited the extent of dialogue between CE teams and MFA, Sida and Embassies.”</p>	<p>Learning from past experience, budgets for country evaluations have increased gradually. A section of costing estimates should be included in the CE planning memo. Advance estimates of likely costs of activities at the time of commissioning should be included when relevant.</p>
<p>Coordination and dialogue with concerned individuals/institutions</p>	<p>There is a risk that Sida and EBA overlap in their evaluations.</p>	<p>“We see it as inevitable that EBA’s future ‘opportunity space’ will be affected by Sida’s plans and vice versa.”</p>	<p>There is an ongoing dialogue between EBA and Sida's evaluation department. The risk of overlap to the extent that it can be seen as double work is considered small.</p>

Observation	Reference to the report	Comments and potential change of methods and procedures
<p>Early dialogue and continued inclusion of various stakeholders in the evaluation process have been limited. Continuous validation and feedback to partner countries have also been limited.</p>	<p>“Limited early and continuing involvement of stakeholders may explain some weaknesses of CE reports in terms of their usability.” “Agenda 2030 make it likely that involvement of countries and other beneficiaries in evaluation will become a stronger expectation in future.” “validation meetings with Embassies did not routinely take place although dissemination meetings in-country were more common. [...] There seemed to be no expectation communicated by EBA that feedback would be provided to partner governments or country stakeholders.”</p>	<p>A continued discussion concerning who should be involved in the study process and how is required. The CE memo should include an internal list of stakeholders that should be consulted throughout the process.</p>
<p>Evaluations have to a greater extent responded against Swedish priorities, rather than included a partner perspective.</p>	<p>“In terms of relevance, evaluations responded to Swedish priorities but less to country priorities, reflecting a move away from Paris Declaration norms over recent years.” “Country involvement in all donor driven evaluations including CEs is considered good practice. This would normally include prior consultation, taking account of country priorities in evaluation design and feeding back evaluation results. This seems to have</p>	<p>The start-up phase of country evaluations should include an analysis of partner countries’ priorities and the relevance of partner country involvement. This should be done in a way that doesn’t overly burden the country's administration.</p>

	Observation	Reference to the report	Comments and potential change of methods and procedures
		<p>happened to a very limited extent in the CE portfolio.”</p> <p>“In terms of coherence CEs were weak in analysing coordination with other donor including multilaterals and the EU and with country plans. In addition, long-term governance also an aspect of aid coordination was not addressed despite being signposted by Statskontoret.”</p> <p>“A consistent weakness of the CE portfolio is the failure to analyse interactions and synergies between Sweden and other donors.”</p>	<p>The terms of reference should, when relevant, include a requirement that an analysis of interactions and synergies with other donors is performed.</p>
<p>EBA’s role in the evaluation process</p>	<p>Challenges during the study process have implied risks related to the study quality which could have been avoided through proactive action from EBA.</p>	<p>“Our interviews suggested that although initially it was accepted in MFA and Sida that access problems were not within the control of the study team, later negative judgements about ‘report quality’ appeared to ignore why this was. More proactive management may have led to greater efficiency in this case; and at the same time mitigate reputational risk for EBA.”</p>	<p>Potential risks should be discussed in start-up meetings. Important deviations between the proposal and the final evaluation (and explanations for them) should be highlighted in the published version.</p>

	Observation	Reference to the report	Comments and potential change of methods and procedures
	<p>EBA has a key role in connecting evaluators with Sida, MFA, and embassies. This is something that has not always worked optimally.</p>	<p>“EBA management are key gatekeepers here, especially by making contacts in MFA and Sida; and via Stockholm headquarters with embassy contacts in-country. Although EBA strengthened these liaison arrangements during CE implementation and since, interviewees agreed that these arrangements were not optimal at the time.”</p>	<p>Consultations in the start-up phase will facilitate continued contact with relevant stakeholders. Initial joint author-secretariat assessment of authors’ needs of contacts as well as access to data.</p>
Evaluation competence and methods	<p>Evaluation teams have often had a greater practical experience rather than methodological expertise. The methods used have not always been adapted to the questions in the evaluations.</p>	<p>“CEs also sometimes aspired to greater rigour and would have benefited from more methodological expertise in evaluation teams.” “additional methods are needed to answer the kinds of questions EBA posed for CEs” “This suggests that preferred skills are more likely to be of a qualitative kind – Theory of Change, ‘theory-based’ impact evaluations, Contribution Analysis; or of the ‘quali-quant’ kind, such as QCA or Process Tracing” “CEs do sometimes use the language of ‘contribution’ but they do not employ methods that would be needed to demonstrate causal and contributory pathways.”</p>	<p>Rigour in design, methodology and methods used is an assessment criterion in the tendering process. EBA strives for a purposeful combination of relevant expertise (methodological, contextual, sectoral) in teams contracted.</p>

	Observation	Reference to the report	Comments and potential change of methods and procedures
Communication, Learning, and Use	It has not always been clear for everyone (those involved and the intended users) what the purpose and intended use of the evaluation has been.	<p>“The teams selected for these CEs with one exception were practice led rather than methodology led. Several CEs had weak descriptions and discussions of their methodological approach”</p> <p>Types of use: Enlightenment or conceptual, instrumental, symbolic, process. “Distinguishing clearly between those CEs that are expected to support instrumental use and those that are to have longer term or more diffuse use and learning purposes would be helpful.”</p>	<p>The report presents different types of use. Internal discussions around the individual and general purpose, aim and intended use of CEs and other studies are internalised in EBA’s working methods. This will facilitate the communication around the purpose of the evaluations.</p>
	Potential users of the study have not always been involved in the study process.	<p>“Opportunities to steer CEs once commissioned in directions that would maximise policy usefulness would need management and Reference Group interventions.” “At the time that the CE portfolio were underway active participation of Sida and MFA representatives in RGs was not mandatory and did not always occur.</p>	<p>At least one person from Sida/MFA/embassy should always be included in the reference group for country evaluations.</p>

Observation	Reference to the report	Comments and potential change of methods and procedures
<p>The communication of results and promotion of use has mainly been focused on the dissemination of the final study.</p>	<p>“Engaging potential users from the outset is likely to encourage all varieties of evaluation use.”</p> <p>“If the aim is that knowledge and insights continue to be available to inform future thought and action, a more extended and interactive communication strategy would be needed.”</p> <p>“In its current form EBA’s strategy to promote ‘use’ is concentrated at two moments. First when planning evaluations by identifying stakeholder in Sida and MFA [...] Second when delivering evaluation reports, [...] The emphasis appears to be to ‘get the message out’.”</p>	<p>As part of a more interactive communication strategy, with the aim to increase learning and use, more interactions with intended users now take place throughout the evaluation process.</p>
<p>Evaluation results need to be “translated” to a policy context to the more useful.</p>	<p>“CEs are reasonably accessible in terms of clarity but would need mediation and translation into policy/practice settings to maximise their usability. Mediation includes well-tried approaches such as preparing short mini-reports or ‘policy briefs; running workshops for anticipated target-groups; and relying on those in ‘translation’ roles such as an</p>	<p>EBA will continue to use existing effective ways to communicate in various forms (such as one-pagers, small-audience and one-to-one presentations, workshops, podcast) and endeavours to find new ways.</p>

Observation	Reference to the report	Comments and potential change of methods and procedures
<p>Quality</p> <p>There has not been a systematic quality assurance of the reports.</p>	<p>research, policy or evaluation unit in a ministry or agency, able to match lessons to their own context.”</p> <p>“For CEs it was only when a final report was submitted that there were formal opportunities for RG and EBA members to comment on report quality. A quality assurance system would require earlier and continuous interactions starting from expectations set at the Commissioning/ToR stage.”</p> <p>“As already noted there was a lack of systematic quality assessments of CEs other than at the endpoint when draft final reports were prepared.”</p> <p>“making explicit the quality criteria against which work will be assessed would be desirable as this would provide evaluation teams with clearer expectations and encourage them to take greater responsibility for their own quality assurance”</p>	<p>The EBA quality policy describes the process for quality assurance. In addition to an expanded work with the policy, there are no plans for change in the process for quality assurance of reports.</p>

	Observation	Reference to the report	Comments and potential change of methods and procedures
	The studies have not always responded to all parts in the terms of reference.	“Although CEs generally answered evaluation questions, they did not deliver on their more ambitious promises to develop or validate a more generalisable model for country evaluations.”	Potential risks should be discussed in start-up meetings. Important deviations between the proposal and the final evaluation (and explanations for them) should be highlighted in the published version.
Ethical considerations	Ethical requirements	<p>“no ‘ethical codes’ were in place for teams or EBA managers to refer to in cases such as this. Being explicit about ethical expectations in advance is common nowadays in evaluation practice. For example, in international evaluations adherence to UNEG ethical guidelines is often a contractual requirement. Furthermore, there was no standard requirement for reports to report on ethical dilemmas and decisions.”</p> <p>“Some policies vigorously supported in the 1970s might be judged very differently in the early 21st century.” “Ethics through the rear-view mirror merits further discussion.”</p>	Since 2017, EBA contracts include formal reference to ethical guidelines. Since 2021 formal reference is made in the contract to the EBA quality policy that includes questions regarding ethics. A reference to the policy and the ethical aspects should also be included in the study terms of reference.

	Observation	Reference to the report	Comments and potential change of methods and procedures
Compilation of results and lessons from previous country evaluations.	There is potential for additional learning from the EBA country evaluations.	<p>“For EBA to gain the maximum from what has been invested in CEs so far, would require a systematic identification of what can begin to be accumulated. This could for example include findings on substantive topics such as capacity-building and sustainability; or strategies when working in difficult contexts for democracy and human rights.”</p> <p>“CE reports contain a wealth of description and discussion of dialogue and other attempts to partner with and influence governments. This material is not systematically organised or within a comparable framework. This too is a theme that could be reviewed and synthesised across all CEs in this portfolio.”</p>	Various compilations of lessons and results from country evaluations could contribute to improved learning related to the results of international development cooperation.