EBA

02

2021

# OBJECTIVES AND MECHANISMS: WHAT DO EVALUATIONS SAY ABOUT THE EFFECTIVENESS OF SWEDISH DEVELOPMENT INTERVENTIONS?

Markus Burman

# Objectives and Mechanisms: What do Evaluations Say About the Effectiveness of Swedish Development Interventions?

Markus Burman

Rapport 2021:02
to
the Expert Group for Aid Studies (EBA)

**Markus Burman** is a programme manager at the Expert Group for Aid Studies and previously worked as a consultant.

**Acknowledgements**

I would like to express my great thanks to the study's reference group consisting of Kim Forss, Johanna Lindström (FCG Sweden), Sven Olander (Sida) and Lennart Peck (MFA) for valuable comments. I would also like to thank Jan Pettersson and Joakim Molander for important comments on the report.

Please refer to this report as: Burman, M. (2021), *Objectives and Mechanisms: What do Evaluations Say About the Effectiveness of Swedish Development Interventions?*, Report 2021:02, Expert Group for Aid Studies (EBA), Sweden.

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



ISBN 978-91-88143-79-2
Printed by Elanders Sverige AB
Stockholm 2022

Cover design by Julia Demchenko

# Contents

# Foreword by EBA

The remit of the EBA is to contribute knowledge as a basis for the development of aid. One central part of this is the assessment of Sweden's official development assistance in terms of its relevance, effectiveness, impact, sustainability, cost-effectiveness and other qualities. Since the EBA does not usually evaluate individual projects or programmes, these assessments are often made at an overarching policy level.

This study starts from the broad question of whether Swedish development assistance reaches its goals. The sample analysed is Sida's decentralised evaluations in the period 2017–2019. A previous EBA study estimates that these evaluations make up 19 % of all evaluations of Sweden's bilateral development assistance, so the sample is not comprehensive. Nor is it the case that all interventions are evaluated. Thus, conclusions are not necessarily valid outside Sida's decentralised evaluations during this period. In addition, this study does not claim to say anything about the evaluations as a whole, only about the parts concerning effectiveness and impact.

Given this, the author of this report finds that Sida's decentralised evaluations often assess effectiveness as good. One possible conclusion of this is that Swedish development assistance generally reaches its objectives. Another possible conclusion is that less successful projects are not evaluated. A further possibility is that the statements made about effectiveness are not reliable. The author takes the view that the interventions may very well have reached their objectives, but that, in general, this has not been substantiated in the evaluations studied.

The question of goal attainment is often divided up into two parts: first, the objectives to be achieved directly by the intervention and, second, more overarching objectives, more long-term transformative objectives (do the activities lead to any form of development?). The author takes the view that there are (at least) two challenges in the evaluation of goal attainment:

(i) Even evaluations of the objectives to be reached directly must be able to demonstrate that observed results have been caused by the intervention, i.e. a causal analysis is essential. The author notes that this causal analysis is not carried out in the evaluations studied. This means that conclusions, which obviously do not need to be wrong, have deficiencies in terms of reliability.

(ii) A full evaluation of the goal attainment of an intervention requires a (causal) analysis of its impacts (long-term, transformative, effects). Without an assessment at the level of overarching goals, it cannot be claimed that the interventions' objectives have been evaluated. Evaluation against overarching goals is essential to be able to claim credibly that aid makes a meaningful difference.

Reliable reporting of results to the Riksdag (Swedish Parliament), the Government and the public requires that overarching goals are evaluated to a much greater degree than is currently done. The author takes the view that Sida's decentralised evaluations are suitable for this type of assessment.

On the basis of these conclusions the author concludes by discussing possible causes of this lack of reliability and potential adjustments of these evaluation activities.

The EBA hopes that this report will contribute to a thorough and much-needed discussion of goal-based evaluation in aid. The report is mainly aimed at people working at an overarching or strategic level on aid evaluation, people carrying out evaluations of Swedish aid and people working on aid effectiveness. We also believe that this report may find users among people working on reporting results at the MFA, Sida and elsewhere and among an interested public.

This report has been written by a person employed at EBA. However, the process is the same as for other EBA reports. The work has been followed by a reference group led by Kim Forss.

The analysis, conclusions and recommendations are those of the author alone.

Gothenburg, 1 July 2021

Helena Lindholm

# Summary

The purpose of this study is to examine what recent evaluations say – and can say – about effectiveness of Sida's aid interventions, and therefore about Swedish aid. This is done of the basis of the definition of effectiveness used by Sida with reference to the OECD DAC criteria for the evaluation of development assistance.[1]

There is extensive literature about goals, goal attainment and evaluation. Parts of this literature are normative and prescribe how objectives and results, management and decisions should be described, designed and evaluated. Another, empirically based part of the literature is consistently more sceptical and critical and gives cause not to have high expectations of analyses of effectiveness in practice.

My review of *previous syntheses* in the area of development cooperation finds that, in the past, Swedish decentralised aid evaluations commissioned by Sida have generally drawn the conclusion that effectiveness has been good in the short or medium term, even though there has generally not been an assessment against overarching goals.[2]

---

[1] The study starts from the effectiveness criterion, but a conclusion drawn during work on the study is that it is probably not desirable or reasonable to strictly separate the effectiveness criterion from the impact criterion regarding the expected results of interventions. An exhaustive and useful assessment of effectiveness has seldom been made without also an assessment of impact, and analyses of both outcomes and impact have to deal with causality. There is also very strong overlap between the criteria, see annex 4.

[2] Decentralised evaluations are commissioned by embassies or at Sida's departments in Stockholm, and generally concern individual projects or programmes. The strategic evaluations decided on by the Agency's Director-General usually take a broader approach, do not focus on individual interventions, and aim to achieve learning at agency-wide level.

Development interventions are often considered to have unclear formulations of objectives, which may have hampered the possibility of evaluating effectiveness through effective operationalisation. In external quality assessments, these analyses have also often been considered to have methodological shortcomings regarding, for example, the causal analysis required to be able to link the results to the intervention.

This new review of 80 decentralised and 6 strategic evaluations demonstrates great pluralism and heterogeneity regarding the overarching goals of the aid interventions evaluated. The evaluators are often critical in one way or another about how objectives/expected results have been formulated in terms of ambition, degree of specification, measurability, realism, consistency and acceptance among the parties involved or time perspectives. However, the lack of clarity in the objectives formulated identified by evaluators has seldom led to a reconstruction of clarification according to the methods available for doing so.

Unexpected (positive or negative) results and effects are asked for in around half the evaluations, despite these being needed to avoid 'tunnel vision' in the assessment of results.

The evaluations in the sample almost only report good effectiveness in the interventions. The most common conclusion is that, so far, the objectives/expected results have mainly been fulfilled. These assessments relate to results as far as the evaluators are able to capture them, primarily outputs and outcomes, even though many evaluations, in line with their terms of reference, also assess overarching goals. However, this is often done partially and incompletely. This means that I cannot give a full picture of the fulfilment of objectives and effectiveness.

Two factors that may have reduced the reproducibility of my review are that the effectiveness assessments in the evaluations are often fairly indirect, multifaceted or unclear, and that overall concerted assessments are not always carried out. A categorisation could not be determined for 10 % of the sample.

Another overall conclusion of the study is that the effectiveness assessments in Sida's decentralised evaluations are probably *not* reliable, particularly because the evaluators have rarely succeeded in showing that the particular intervention has contributed towards or achieved a change observed. The evaluators rarely use new qualitative or quantitative evaluation methodology. The methodology for answering the question of effectiveness is not described in a transparent way, and is rarely properly adapted to the specific evaluation question.

Relatively few of the evaluations' terms of reference ask for analyses of mechanisms or underlying explanations of effectiveness/results achieved. This is a central issue if the purpose of the evaluations is learning/utility (Stern et al, 2012).

A small number of the evaluations in the sample find weak effectiveness. A separate analysis of the learning potential of these studies shows that they provide conceivable partial explanations of weak effectiveness linked to the management, design, organisation and monitoring of the interventions. However, little is said about external explanations or mechanisms. The evaluators list possible factors, barriers or measures that they believe can improve the intervention, rather than analysing how or why the results have or have not been achieved. Overall, this limits the potential for learning.

The potential for learning in the evaluations also decreases because of their similarity in conclusions. Thus, they cannot contribute new knowledge, from an overall agency-wide perspective, about what works where, when, how and why. If almost all reports conclude that the interventions achieve their goals, then the variation of experience that is important for learning is nowhere to be found.

The conclusions and analyses of the report are summed up in the form of two main conclusions, four recommendations and five questions for future analysis:

## Two main conclusions

1. The evaluations in the sample usually draw the conclusion that the intervention (mainly) achieves its objectives. This assessment is mainly made at the output and outcome levels (chapter 4).
2. In general, their conclusions about effectiveness are probably not reliable (chapter 5).

## Four recommendations

1. The quality of the decentralised evaluations regarding evaluation design and methodology for analysing effectiveness and expected results should be strengthened considerably (see chapter 8).
2. More Swedish development interventions can and should be evaluated against overall, long-term goals. This also appears to be in demand from Sida staff commissioning evaluations (chapters 3 and 4). Today Swedish development assistance is highly focused on core support and regional and global forms of support. Such interventions are hard to evaluate. Sida should therefore analyse the evaluability of these interventions in terms of the conditions for evaluating and assessing the goal attainment and effectiveness of Swedish development assistance in the future (cf. the discussion in chapter 5).
3. Assessments of expected results should be supplemented to a greater extent by analyses of unexpected results and possible side-effects in development assistance (see chapter 3).
4. In the future the evaluations should focus more on goal attainment and on trying to understand how and why interventions do (not) reach their objectives (see chapters 2, 3 and 7 and annex 4).

## Five questions for future analysis

The are a large number of adjacent and closely related questions about Swedish aid, evaluation, monitoring, reporting and effectiveness that came up in the course of this work but that have not been given space in this study. The following five questions should be analysed and illustrated particularly in the future.

1. How is effectiveness assessed more generally in Sweden's bilateral aid? Other than evaluations, what supporting information do Sida, the Folke Bernadotte Academy (FBA), Swedfund and others use to assess and report effectiveness to the Riksdag and the Government? How reliable are this supporting information and these assessments?
2. How can and should Sida's information on results (evaluation, monitoring and reporting) be aggregated and summed up at strategy and country level?
3. How can Sida and other aid donors assess unforeseen results and side effects in a more comprehensive, systematic and reliable way? How is this evaluated today? How can and should this be done?
4. How reliable are the conclusions in Sida's strategic and partner-led evaluations?
5. How can the challenging causality assessments in evaluations be developed to achieve the 'good enough' level also in evaluations of small or limited development interventions, or where the evaluation budget is limited?

The final chapter of the report discusses four areas as possible points of departure in an explanatory analysis of the reliability of Sida's decentralised evaluations:

- Demand, needs and incentives.
- The system for procurement, contractring, feedback and following up on evaluation services.
- Supply in the evaluation market.
- Knowledge about evaluation and about learning and development in operations.

# 1. Why this study?

Questions concerning effectiveness/the fulfilment of objectives and results are among those discussed most and for longest in development assistance. All government agencies in Sweden are required to conduct their activities effectively.[3] For international development assistance the Government's Policy framework for Swedish development cooperation (Government, 2016) lays down that:

> "Information on the effects of development cooperation is important for learning and accountability. This is contingent upon successfully functioning monitoring and evaluation systems in Sweden and in partner countries and international organisations."

Even though research has shown that interest in results in Swedish development assistance has gone in waves (Vähämäki, 2017), the Government, the Riksdag, Sida's officers and the public need to get reliable information about what these activities achieve. The ability to distinguish success from failure is of central importance for learning, adaptation and accountability in democratic systems of government. The government-appointed Analysis and Evaluation Inquiry (SOU 2018:79) took the view that the Government mainly asks for knowledge about impacts, efficiency and, specifically, effectiveness from its analysis agencies.

In this report I study the evaluation of effectiveness in a sample of Sida's evaluations. Evaluation can be considered to have a special standing among supporting information and monitoring processes

[3] See the Budget Act, the Government Agencies Ordinance and the Government's public administration policy. The Swedish National Financial Management Authority (ESV, 2006) writes that "[m]easures taken to reach given objectives must be measures that actually lead to the fulfilment of the objectives, and not measures that are only aimed at fulfilling the objectives".

by virtue of high expected reliability, the potential for learning, accountability and independent assessment. Evaluations are particularly important in areas where the result of the activities cannot be observed directly by the funder, such as the area of aid.[4] The Government, the Riksdag and other actors are therefore dependent on Sida's and EBA's evaluation work.

The purpose of this study is to examine *what recent evaluations say about the effectiveness of Swedish aid interventions and about Swedish aid*. The starting point for this study is the definition of *effectiveness* given by OECD/DAC for aid evaluation.

**Definition of effectiveness in the study**

"The extent to which a development intervention's objectives were achieved, or are expected to be achieved, taking into account their relative importance."[5]

My interpretation of effectiveness, called the goal attainment model, is a standard interpretation that combines the DAC criterion with the established understanding of effectiveness in the field of evaluation according to, for example, Eliasson (2006), Vedung (2012) and Faugert and Sandberg (2012).

---

[4] "One important difference between international development assistance and other publicly financed activities is that aid is implemented in other countries. This means that recipients of aid do not have direct paths to influencing decisions about how money is distributed and used. Users of aid do not have the right to vote in Swedish elections, do not have the right to appeal the decisions of Swedish government agencies and they have small possibilities of taking part in the public debate in Sweden. Swedish taxpayers have, on the other hand, limited possibilities of following how aid funds are used and what result they lead to. In other words: both learning and accountability are made more difficult by geographical distance and the organisation of the activities" (Burman and Hårsmar, 2015).

[5] OECD/DAC's evaluation criteria were revised in 2019. The new definition is: "The extent to which the intervention achieved, or is expected to achieve, its objectives, and its results, including any differential results across groups" (OECD/DAC, 2019). The evaluations analysed were commissioned when the previous definition still applied.

The starting point for this study is the OECD/DAC effectiveness criterion. One conclusion that has been drawn is that it is, in practice, is not desirable to strictly separate the *effectiveness* criterion from the *impact* criterion regarding expected results. An exhaustive assessment of effectiveness has seldom been conducted without an assessment of impact, and causality has to be addressed in analyses of both outcome and impact. Annex 4 discusses the definition of goal attainment and how the criteria overlap.

The two overall questions in this study are:

1. *What conclusions are drawn about effectiveness (or expected results achieved) in recent evaluations of Sweden's bilateral aid interventions.*

Do the conclusions differ depending on the form of the intervention (project, programme, organisation support) or its channel (civil society, public administration, companies), geography (region, low- or middle-income countries) or other context (long-term development cooperation or fragile or conflict affected states)?

2. *Are the conclusions about effectiveness reliable and, if so, useful for learning?*

Do the evaluations make clear how the question of effectiveness has been answered and what limitations the methods chosen have? Is there an analysis of how the intervention has affected effectiveness specifically in relation to other influencing factors. Is it reasonable to evaluate effectiveness at the point in time when evaluations are conducted? What factors are said to have affected effectiveness as a basis for learning?

With the guidance of an analytical framework (annex 2) these questions have been answered through a meta-evaluation of 80 decentralised evaluations from the period 2017 to 2019 ordered by Sida in Stockholm or by embassies. The evaluations generally concern single interventions or organisations. In addition, six centrally conducted ("strategic") evaluations in the period 2015–2019 are included.[6] These study broader themes or recurring modalities.

One important restriction of the scope of this study is that I have not analysed Sida's evaluation activities as a whole but have studied the assessment of effectiveness made in single, mainly decentralised, evaluations. Even though almost all (93 %) of the decentralised evaluations in the period selected are included, they are not representative of Sida's aid portfolio as a whole. Sida's evaluations can also change over time. The sample can therefore not be viewed as representative of either Swedish aid or Sida's evaluations other than the years included in the study. There is a thorough discussion of the study's method, sample and representativeness in annex 1.

Nor do I claim to give a comprehensive picture of individual evaluations since they also have other evaluation questions to answer (about relevance, efficiency, sustainability, etc.).

The most important limitation of the study is the reliability of the evaluations analysed, which clearly affects what conclusions can be drawn about the effectiveness of the interventions. However, this fact is made a subject of study in its own right (chapter 5).

The study is aimed primarily at those working with the governance of Swedish aid (MFA, Riksdag); Sida's senior management; and persons ordering or conducting evaluations of Swedish development assistance. Other important target groups are persons working with evaluations, assessments of results and result-based management in Swedish aid. Another important target group is members of the public in Sweden who are interested in the results of development assistance.

Goal-based evaluation is discussed below on the basis of a number of sources in the *literature* (chapter 2). Then (in chapter 3) a mapping is made of goals and objectives in the projects studied by the

---

[6] These are now referred to as central evaluations by Sida.

evaluations. What are the goals or expected results being evaluated? What are Sida's commissions like? How is the scope of the question of effectiveness defined and how does it vary? Chapters 4 and 5 deal with the *study's two main questions* about the conclusions and reliability of the evaluations, and chapter 6 presents the conclusions of *previous synthetic studies* of effectiveness and the evaluation of effectiveness in Swedish aid. Chapter 7 analyses the evaluations in the sample that report weak effectiveness and focuses on *explanations and underlying factors* as a basis for learning. Chapter 8 presents two main conclusions and four recommendations. It also proposes five issues that should be studied in *future studies*. Chapter 9 sets out a number of concluding reflections about reliability. All quotes translated in the report have been translated by the author.

# 2. What is goal-based evaluation?

Goal-based evaluation is a very common form of evaluation and is seen by some as so fundamental that they equate evaluation with assessment against the objectives of an intervention (see Nachmias, quoted in Vedung, 2012).[7] The model has been considered to work "as a norm for what evaluation is and how evaluation should be carried out" (Sandberg and Faugert, 2012).[8]

A common definition is that a *goal* is a "description of a desirable future state" (Krogstrup, 2017), or "descriptions of what an object will look like or be a some point in the future" (Lindgren, 2012). That object is then "the person, population, organisation, process or something else that needs to be changed and that a given intervention is aimed at". Goal-based evaluations analyse whether an intervention has reached or contributes to expected or planned results.

Evaluators who study effectiveness start from the statements of objectives written down by those responsible or reconstructed by the evaluators. One common view is that – to be usable in monitoring and evaluation – objectives should be formulated in a specific way, for instance according to the SMART criteria (Doran, 1981) stating that objectives should be Specific, Measurable, Agreed, Realistic and Time-bound.

Vedung (2012) underlines that, in practice, an analysis of effectiveness asks two questions: "1) To what degree have outcome-goals been achieved (the question of effectiveness)? 2) Has the intervention contributed to goal achievement and, if so, by how

---

[7]A definition of that kind is controversial. In aid evaluation the starting point is often OECD/DAC's six evaluation criteria (see OECD/DAC, 2019).

[8] There are several alternatives to the goal attinment model. Examples include evaluation using assessment matrices, Rubrics-Enhanced Evaluation (Davidson, 2004), principles-focused evaluation (Patton, 2017), indicator-based evaluation and goal-free evaluation (Scriven, 1991).

much (the question of causality)? Thus, it is not possible to evaluate the effectiveness of an intervention without addressing or reflecting around the question of causality. One common misunderstanding starts from the logical framework (figure 1) and assumes that causality need only be addressed at the impact stage.

The Swedish Financial Management Authority (ESV, 2006, p. 7) distinguishes between *descriptive* and *explanatory* analysis of goals and points to the need to determine that it is actually the intervention that has contributed to a goal being achieved or influenced. This applies not least to the stages of outcome and impact in the logical framework.[9]

**Figure 1: The logical framework**

1. Input → 2. Activity → 3. Output → 4. Outcome → 5. Impact

Evaluations that claim to assess effectiveness but do not take account of causality or external influencing factors risk being wholly or partly misleading. Demonstrating the intervention's own contribution to the result is crucial to making it possible to determine the value and future direction of the intervention (Mayne, 2001). The challenge in assessing causality and influence from a number of factors in relation to a specific objective or outcome increases with the passage of time and the degree of complexity (Forss et al. 2011). However, the development of methods for handling causality in evaluations using qualitative and quantitative methods has been rapid in recent years, and the number of design alternatives is increasing all the time (Stern et al, 2012; Sandahl and Petersson, 2016; Molander, 2017).

[9] Causality in output-based objectives can, like input and activity, partly be overviewed in the activity being evaluated, by target groups and indirectly by the evaluator, but the additionality of external support and its influence on short-term activities and on output are not obvious. There are many examples of public interventions with weak additionality at the input and output levels.

The critical literature on goal-based evaluation is extensive. Table 1 summarises some of the strengths and weaknesses highlighted in different parts of the literature about the goal attainment model.

**Table 1: Strengths and weaknesses of the goal attainment model**

| | Strengths (+) |
|---|---|
| Democracy | The model affirms the view that policy objectives often play a central role in democratic systems. Taxpayers have a democratic right to know whether public interventions live up to publicly declared intentions. The model can take official political objectives seriously. |
| Objectivity | The model can give an impartial, objective solution to evaluation as a normative assessment activity. |
| Simplicity and transparency | Relatively simple to apply and communicate to decision-makers, stakeholders and the public. |
| | **Weaknesses/challenges (-)** |
| Airy, unclear or simplified | The objectives of public interventions are often so unclear, weakly prioritised and numerous that they are useless as instruments for assessments. |
| | Complex operations may have difficulty in simplifying their activities into a few objectives. Phenomena that are difficult to measure risk being given no priority. |
| Inflation and hidden motives | Objectives are often set to generate sympathy or give legitimacy and not because they are to be achieved or form the basis for evaluation. |
| | The model can be blind to the role that unreported motives play in the formulation of objectives and in public interventions. The “real” objectives are not always formulated or communicated. |

| | |
|---|---|
| Unforeseen side effects | With a focus on foreseen formal objectives, an evaluation may neglect that it is not unusual for public interventions to have strong, unforeseen (negative or positive) side effects. |
| | **Weaknesses/challenges (-)** |
| Top-down management, rigidity, tunnel vision and sub-optimisation | Management by objectives risks creating top-down managed, rigid, sub-optimised activities that do not see the need for learning and adaptation and that focus instead of shorter term results that are easy to measure or manipulate. Evaluations of effectiveness can then be interpreted as a part of such a harmful system. |

Source: This summary is based on Vedung (2012), Schutt (2016) and Alexius (2017).

Many evaluators would probably agree that airy, unclear and simplified objectives do at least present a challenge, even though they need not be problematic in practice. Thinking about evaluation is based in part on the ability to initially and continuously specify thorough descriptions of what is to be achieved and how something is to be achieved, with careful use of concepts in the operationalisation of evaluation questions (see Leeuw, 2003, on the reconstruction of objectives and theories of change in evaluation). Experienced evaluators are expected to be used to dealing with imprecise objectives in dialogue with commissioners and those responsible for interventions.

At the same time, it is easier to assess effectiveness in interventions with a clear, precise and suitably ambitious formulation of objectives. Operationalisation is made easier and the conditions for valid, reliable conclusions are enhanced.

The objection about the "real" or tacit objectives and intentions is also apt. They can be difficult for the evaluator to handle since there may be incentives for commissioners and those responsible for interventions to actively steer away from the question when an evaluation is to be carried out.

The Swedish Financial Management Authority has highlighted the fact that objectives in central government activities do not need to live up to the SMART criteria if those responsible have a clear idea about how to monitor or evaluate the objective (Eliasson, 2006). Unclear objectives can be made more precise through operationalisation. This requires a process for reconstructing, along with stakeholders and those responsible, what is to achieved and operationalising these expectations as indicators, metrics or criteria for success.

One possible complication is also due to the distinction between objectives to aim for and objectives to achieve.[10] The assessment of effectiveness differs if the objective expresses a long-term visionary endeavour or if the idea is that the objectives will really be reached. If a project portfolio contains both interventions with objectives to aim for and objectives to achieve, it can be doubtful whether effectiveness can be compared between the interventions.

Easterly (2006) has suspect international aid of *Utopian Planning*. He argues that the objectives are unrealistic relative the inputs available to aid. The ambitious objectives generate unachievable expectations and disappointment while making evaluation more difficult through a "missing middle" between overall objectives and intervention objectives. The individual interventions only aim *indirectly* at achieving the objectives specified by the political level. However, the Government objectives, as expressed in various development assistance strategies, differ in this respect.[11]

---

[10] This distinction is used in school curriculums, for example. One example of an objective to aim for is the zero vision in traffic planning.

[11] The strategy for support through civil society in 2016–2022 is, for example, expected to contribute to achieving "stronger capacity among civil society actors in developing countries" and a "more favourable societal climate for civil society organisations in developing countries" The strategy for Sweden's global development in sustainable economic development in 2018–2022 shall, for instance, achieve "more effective domestic resource mobilisation", "greater financial stability" and "action to counter corruption". There can be great differences in ambitions within and between strategies.

In the discussion on goal free-evaluation (Schriven, 1991) it has been said that goal-based evaluation can lead to tunnel vision. Advocates of the model, which is seldom applied, argue that evaluations should focus on what interventions actually do and achieve. It is not the intentions that count but what is achieved.

## Evaluability and drawing conclusions

Karlsson Westman (2011) highlights the fact that the choice of effectiveness as an evaluation criterion is made on the assumption that "there is a link between objectives and results that can be measured", i.e. an assumption of evaluability, that effectiveness can be assessed at the point in time when the evaluation is carried out, that relevant data are available or can be collected, that the evaluation team has the ability and will to make a reliable operationalisation of the objectives and conduct the necessary causal analysis.[12]

One challenge in the evaluation of effectiveness is often arriving at a clear value judgment of the intervention being evaluated (Forss, 2007). These value judgments may be important if a donor is, for example, going to decide whether to scale up an intervention or to end the support. Davidson (2014) argues that it is not sufficient to draw conclusions about whether or not the objectives have been achieved; the evaluator should also clearly appraise effectiveness. Not doing so limits the added value of evaluation and usefulness for decision-makers.

In many aid interventions there are several objectives to evaluate so there is no obvious way of aggregating the analysis of effectiveness (for example in a situation where the analysis is judged to be good for two objectives but weak for two others). The implementer and the funder of the project my give priority to different objectives and

---

[12] Holvoet et al (2018) take the view in a study of evaluability in Belgian development assistance that "the majority of the projects have conditions in place for evaluating internal and external effectiveness in a satisfactory way".

there may be shifts of objectives or reprioritisation over time. A closely related difficulty is linking up effectiveness at project level with overall objectives at strategy level or for the policy as a whole (cf. Christoplos et al, 2014).

It should be pointed out that an evaluation that only assesses short-term output, outcomes, sub-objectives or activity objectives rather than expected ultimate objectives has *by definition* not carried out a full assessment of effectiveness. An assessment of effectiveness is made relative *expected* ultimate objectives/results for an activity.

Goal attainment is also limited if efficiency is not included, i.e. if the costs of the intervention are not taken into account in the assessment of effectiveness (Sandberg and Faugert 2012). Good effectiveness does not have the same value independent of costs. Aid evaluations generally try to achieve a more comprehensive assessment by using all or several of the OECD/DAC evaluation criteria.

## Short-term change, expected and unexpected results

Good effectiveness regarding shorter term performance in a project does not guarantee good results in the medium term and good results in the medium term do not guarantee good long-term impacts either. On the contrary, short-term changes often say relatively little about results and impact in the long term. Forss (2021, p. 200) notes that:

> "there is a risk that short-term evaluations miss out on changes that are transformative, that occur at tipping points, and that are episodic and punctuational, exponential and radical-core".

In development assistance the OECD/DAC evaluation criteria place unexpected, positive and negative, results and side effects under the Impact criterion. However, it is of central importance to also discuss and consider the expected in relation to the unexpected in assessments of effectiveness (see also the discussion in annex 4):

> "If the evaluation is concentrated, as in the goal-attainment model, on the question of whether the reform's own objectives have been realised, it will, by definition, not trace serendipities and unforeseen side effects. The reason is that objectives may not have been set for them and that they are therefore not included in the field examined since they are unforeseen. This creates blind spots that can give a distorted picture of what the reforms have achieved" (Vedung 2012, p. 101).

## Flexibility and adaptation

One discussed weakness of the effectiveness model that can be questioned is top-down management and rigidity (see table 1). In recent years researchers have highlighted the need for flexibility, realism, learning and adaptability in aid interventions (for example as part of the Problem Driven Iterative Approach (Andrews et al, 2015) and Doing Development Differently (Honig & Gulrajani, 2017).

An opposition can arise between goal-based evaluation and values like flexibility, realism, learning or adaptation if the evaluator does not take account of the need for flexibility and adaptation regarding how ambitions and statements of objectives relate to one another. But if an intervention at a particular point in time does not achieve its objectives, then it – quite simply – does not achieve its objectives. Criticism of goal-based evaluation from an "adaptive" perspective can therefore sometimes be based on a confusion of goals and means. As concluded by Vedung (2012), there is nothing in the goal attainament model that says that interventions give rise to intended results. On the contrary, that is precisely what is being examined. Given that objectives need to be adjusted and the level of ambition lowered or raised to establish the right balance between challenge and realism, an inflexibly conducted evaluation of objectives may nevertheless clash with an adaptive approach. Initially, a clear purpose of results-based management was flexibility and delegated implementation (Jacobsson and Sundström, 2001). Nor should changes be made to the strategy, ways of working or process in an activity whose effectiveness and results are not known.[13]

## Incentives and power

The perhaps most famous objection to goal-based evaluation and management is called *Goodhart's Law* (Goodhart, 1975), originally formulated for monetary policy, which states that

> "Any observed statistical regularity will tend to collapse once pressure is placed upon it for control purposes".

Goals may perhaps be unclear not because they cannot be formulated correctly, but because there is no wish to do so.[14] The incentives to draft clear objectives may be weak for several

---

[13] It is more troublesome from an adaptive perspective when the logical framework (e,g, LFA, the Logical framework Approach) is used as a governing project management tool and then steers both objectives and means at the same time. LFA matrices therefore entail stronger interventionist management than traditional results-based management.

[14] Tarchys (2006) takes the view that: "report after report comes to the same monotonous conclusions: Decision-makers must clearly specify what they want to achieve. But decision-makers do not want to do so. For several different reasons they are not all that inclined to be more precise. One reason may be that they do not want to make themselves vulnerable by setting clear objectives that then are not achieved; another may be that vagueness and ambiguity facilitate compromises and unanimous decision-making; a third may be that politicians as a collective do not need to exercise clear leadership in relation to the administration."

stakeholders in, for instance, development interventions. Effectiveness in an activity often has to be reported to several funders, so too clear objectives can be downright impractical.[15]

Pritchett (2002) argues that limited investments in impact evaluation within a policy area can be explained by the incentives for stakeholder groups not to finance credible knowledge since it can undermine the ability to mobilise political and budgetary support. It has also been shown internationally that evaluators sometimes experience pressure to deliver the "right" message to their commissioners (see, for example Pleger et al, 2020).

At the same time, it does not have to be seen as illegitimate for an official to, for example, commission evaluations to find evidence for the effectiveness of a specific type of intervention as part of an internal negotiation about resources for and the direction of the activity. This is given that the person carrying out the evaluation upholds basic quality criteria for evaluations, such as reliability and ethical implementation.

## Effectiveness, learning and use

From a learning perspective, effectiveness evaluation often means, in practice, what is called single-loop learning (Argyris & Schön, 1976). Objectives as such are not always questioned, only effectiveness is.[16] The positive feature of this is the evaluator's democratic subordination to the policy objectives, which enhances the legitimacy of the evaluation. But this is based on the evaluator actually starting from the overall or policy objectives in their analysis,

---

[15] The fact that reviews constantly show that objectives in aid interventions are unclear (see below) indicates that clarification or reconstruction of objectives to make evaluation possible should be done partly by the evaluator, and not solely by the intervention.

[16] In aid, relevance of objectives is often assessed using the Relevance criterion.

which is not always the case. In practice, evaluations often focus on short-term or intermediate project objectives not formulated by the political sphere.

It is of central importance for learning and utility that evaluations assess *why* and *how* a certain goal has been achieved. What mechanisms explain the degree of effectiveness? Why did interventions work or why did they not work? Would they work in different circumstances, in the future, for whom, why, how in that case, and so on (Stern et al, 2012). Have we actually formulated the right objectives? The final question here can be called double-loop learning, as in Argyris (1993). Some researchers consider that there is also something called triple-loop-learning, that is about "learning to learn" and that can, for example, correspond to a process that more fundamentally questions an organisation's capacity for learning through objective-based evaluation (Swieringa and Wierdsma, 1992). Are there, for instance, more fundamental principles, factors or components in evaluation processes that mean that they do not support the organisation in its endeavours to learn?

## Summing-up: literature about the goal attainment model

There is extensive literature about objectives, effectiveness and evaluation. Part of this literature is normative and prescribes how objectives and results should be described and how management and decisions have to be designed and evaluated. Another part of this literature is descriptive, and shows how real life has a tendency to mess things up. The descriptive literature is, on the whole, more sceptical and critical, and is also empirically based. So there can be reason not to have high expectations of an analysis of the evaluation of objectives in practice.

# 3. Objectives and remits in 80 evaluations

This first empirical chapter charts the objectives of the interventions and the expectations on the evaluations in a sample of 80 decentralised evaluations. However, I begin with a short background to Sida's evaluation work, focusing on decentralised evaluations.

Since their appearance in 1991, the OECD/DAC evaluation criteria have become an international standard for assessment criteria in aid evaluation (OECD/DAC, 2019, see also Chianca, 2008). Assessment of the effectiveness of an intervention is included as one of six criteria.

The criterion would probably have been included in the evaluation of Swedish aid interventions even without the OECD/DAC criteria. Some processes and events that may have fostered a focus on goal-based evaluation in Swedish development assistance are evaluation as an "administrative policy megatrend" prescribing evaluations in almost all areas (Vedung, 2003), cyclical result agendas (Vähämäki, 2017) and the international aid effectiveness agenda (Paris, Accra, Busan) that highlighted the importance of results.[17] Sida also conducted evaluations of the effectiveness of aid interventions before the appearance of the OECD/DAC criteria.

---

[17] It is a dubious step to link objective-focused evaluation in a one-sided way to the complex of ideas called *New Public Management* (Hood, 1991). Vedung (2004) takes the view that evaluation activities in Sweden have emerged through "sedimentation" following four historical waves from 1960 and later: a science-focused wave (with roots in the US and the 1960s), a neoliberal wave, a dialogue-oriented wave and an evidence wave. In the aid area Vähämäki (2017) studies four "result initiatives" launched by Sida in 1971, 1981, 1998 and 2012. Evaluation and assessments seem to have been (more or less) important components of all these initiatives (for an early example, see: *Resultatutvärdering – ett programförslag [Result evaluation – a programme proposal*, SIDA 1971). The OECD/DAC criteria have, however, probably contributed strongly to institutionalising effectiveness as an assessment criterion in aid evaluation.

## Evaluation at Sida

Sida applies a utilisation-focused evaluation approach (Patton, 2008). The emphasis is on "identifying who the users are of a specific evaluation […] If it is unclear what the intended use of an evaluation is or who the intended users are, the evaluation will not be carried out" (Sida 2018; 2020). At the same time, Sida states on its website that the purpose of its evaluation work is accountability by providing transparency[18]

The agency distinguishes between three forms of evaluation: decentralised evaluations, strategic evaluations and partner-led evaluations. Decentralised evaluations (which this study focuses on) are commissioned by embassies or by Sida's departments in Stockholm and most often deal with individual projects or programmes. Strategic evaluations, which are decided on by the agency's Director-General, generally take a broader approach and are intended to generate learning at an agency-wide level. Partner-led evaluations are commissioned by Sida's partners but assess interventions funded by Sida.[19]

Sida' s representative in this study's reference group stresses that decentralised evaluations are to be seen as support in dealing with interventions, in specific decision situations or on the basis of special needs. They are often commissioned ahead of a new project phase and are, according to the Evaluation Unit, to be viewed in a context in which they are often carried out half way into the project cycle as a practical, intervention-centred and needs-focused activity. Sida also emphasises that its partners carry out evaluations and that Sida is therefore one of several actors analysing, monitoring and evaluating development assistance. At the same time, Sida makes frequent

[18] https://www.sida.se/Svenska/Samarbetsparter/resurser/utvarderingar/ (Read 2020-06-15).

[19] No partner-led evaluations are included in this study.

reference to evaluations in its annual report (2019) and on its new website, for example, when it communicates its evaluation work in relation to objectives.[20]

Responsibility for the quality of the decentralised evaluations is shared by the project-managing author, the person who commissioned the evaluation and the consultant responsible for quality. Sida's Evaluation Unit is only responsible for the quality of the decentralised evaluations in an institutional sense since it procures, administers and quality assures the framework agreements that form the basis for the evaluations.[21] The Unit follows up how matters are functioning with suppliers, it provides methods support, offers advice, publishes the evaluation handbook, has dialogues about terms of reference, provides training, follows up the framework agreement and discuses conclusions and quality in the Unit's annual report.

Sida bases its evaluation work on the OECD/DAC evaluation criteria and its definition of effectiveness and specifications of quality. Sida's evaluation manual (2020) does not discuss the effectiveness criteria apart from reproducing the OECD/DAC definition. The handbook is clear that evaluations have to be reliable and it stresses the importance of integrity in the evaluator (ibid p. 10).

The following review of the interventions' stated goals and Sida's commissions is based on 80 decentralised evaluations published in 2017–2019 (for more information about the sample, see annex 1).

---

[20] https://www.sida.se/sa-fungerar-bistandet/sa-har-mater-sida-resultat (read 23 March 2021).

[21] Email correspondence , Sven Olander, 11 March 2021. It has emerged in dialogue with Sida's previous chief evaluation officer that Sida's Evaluation Unit was not judged, at that time, to have responsibility for the quality of the decentralised evaluations.

## What objectives are to be reached?

Table 2 lists the stated goals of the interventions regarding overall intervention objectives by frequency. The compilation demonstrates great pluralism regarding what types of long-term objectives the interventions have.

### Table 2: Overall intervention goals in the evaluations, by frequency

| | |
|---|---|
| Change or improve the functioning or ways of working of public institutions or capacity in a sector | 12 |
| Strengthen implementation of or compliance with human rights in one or more areas (thematically or geographically) | 9 |
| Strengthen capacity to conduct research | 8 |
| Strengthen the use of research results | 7 |
| Create mobilisation or participation or generate demand for democracy | 5 |
| Reduced poverty in a region, country or part of a country | 4 |
| Economic empowerment for a particular group | 4 |
| Conflict resolution, peace, peaceful societies and security | 3 |
| Spread knowledge, train, develop skills, inform about an area/issue | 3 |
| New and better jobs/job opportunities | 3 |
| Better living conditions or higher incomes | 3 |
| Strengthen trade | 3 |
| Reduce violence in various forms or contexts or in relation to different groups | 3 |

Source: All decentralised evaluations in the sample 2017–2019. Only goals that occur three or more times are included in the table. The goals have been produced inductively by coding overarching goals in broad categories.

The level of ambition differs greatly between interventions since some are intended to spread information or strengthen knowledge while others are intended to achieve greater economic growth or reduce violence.[22] Interventions are sometimes, but far from always, intended to achieve major ambitious objectives such as poverty reduction.[23] Objectives such as human rights, reduced violence, increased trade, new jobs and the like occur. However, the largest group of interventions have objectives like changing or improving public institutions' ways of working, strengthening capacity, strengthening mobilisation, participation, greater use of knowledge, spreading knowledge and so on.[24]

An issue, not addressed in this study but raised in the reference group, is that the goals are often formulated in terms of something increasing, being strengthened, contributing or expanding, i.e. in the form of a verb and not as a state. It is rare for nothing to be happening in an aid project; the actors keep working, hold seminars or training courses, get organised, arrange workshops, implement activities, etc. To exaggerate a bit, the conclusion could be drawn that it is only when nothing is happening or the activities come to a complete standstill that the goals concerned are not (being) reached.[25]

Of the 68 evaluation that only evaluate one intervention, 27 (40 %) have more than one overall objective. At the same time, the focus is often on short-term outcomes and results. In practice, the evaluator therefore has considerably more than one objective to consider. The

---

[22] A large number of the evaluations had objectives unique to the intervention evaluated The table only includes objectives found in at least three evaluations.
[23] At the same time, the poverty objective may be implicit in the intervention.
[24] The level of ambition regarding objectives in development assistance should be understood in relation to the specific context. What is an ambitious objective at the tax agency in Afghanistan does not need to be seen as ambitious at the tax agency in Tanzania.
[25] If that hypothesis is correct, it means that the goal attainment model does not work because the objectives have not been formulated optimally. Ambitious projects or well-formulated goals of future states-of-the-world risk being disfavoured in comparative analyses. The SMART criteria state that objectives have to be realistic (R), but they also have to be challenging.

effectiveness criterion sometimes contains 10 evaluation questions specified in the terms of reference. In addition to this the evaluator has to answer questions about relevance, efficiency, sustainability, etc.

The impression is that the interventions more often have objectives to attain rather than visionary objectives. In the evaluations there is a recurring expectation that structures of objectives and logical frameworks will act as guides for interventions, monitoring and evaluation. However, it has not been possible to determine what share of the objectives are objectives to attain rather than visionary objectives, which can affect comparability between interventions.

The interventions often have the purpose of strengthening various organisations, such as civil society organisations, universities and public authorities (table 3). A third of the interventions evaluated only expect direct results for poor people, citizens or holders of rights or in nature or ecosystems.[26]

### Table 3: Are the overall long-term results to be achieved direct or in processes, organisations and activities?

| | | |
|---|---|---|
| In both | 29 | (36 %) |
| In processes, organisations and activities that shall be strengthened, built or changed[a] | 27 | (34 %) |
| Directly in the communities affected[b] | 23 | (29 %) |
| Not relevant/cannot be seen | 1 | (1 %) |
| Total | 80 | (100 %) |

Notes: [a]That can then directly/indirectly influence people/society/nature. [b]Directly in relation to citizens, rights holders, in nature or ecosystems. Source: All 80 decentralised evaluations in the sample 2017–2019.

[26] This should not be viewed as an appraisal of the intervention.

## What is to be evaluated?

The analysis of effectiveness is a central part of the remit of all 80 evaluations. However, the review (table 4) shows that Sida often does not explicitly state in the terms of reference that it is the fulfilment of objectives specifically that is meant in the assessment of effectiveness, and that other wording is used instead such as expected/planned results or outcome components.

### Table 4: Do the terms of reference *explicitly* ask for an assessment of goal attainment or are other concepts used?

| | | |
|---|---|---|
| Yes, goal attainment is asked for | 34 | (43 %) |
| No, other concepts are used | 43 | (54 %) |
| No specification of the remit | 3 | (4 %) |
| Total | 80 | (100 %) |

Source: All 80 decentralised evaluations in the sample 2017–2019. All these evaluations focus on the effectiveness criterion.

As shown in table 5, the evaluators often (66 cases) discuss the formulation of the intervention's objectives or expected results to some extent. Their criticism is often about how objectives/expected results have been formulated in terms of ambitions, specification, measurability, realism, consistency, acceptance among the parties involved or time perspectives.

### Table 5: What do the evaluators say about the formulation of objectives/expected results.

| | | |
|---|---|---|
| The evaluators are critical/partly critical | 54 | (68 %) |
| The evaluators are positive/very positive | 12 | (15 %) |
| The formulation is not discussed. | 14 | (18 %) |
| Total | 80 | (100 %) |

Source: All 80 decentralised evaluations in the sample 2017–2019.

The effectiveness criteria corresponds in practice, in the material asked for by Sida, to a large number of related questions and sub-questions in the evaluators' terms of reference (table 6). Many commissioners (79 %) specify particular types of results/objectives or break-downs of results that the evaluator is expected to analyse.[27]

Just over a third of the terms of reference (38 %) ask for analyses of mechanisms or explanations concerning effectiveness or results. Around a quarter of the terms of reference (23 %) want to see analyses of the intervention's result and monitoring work, while a fifth (21 %) ask for assessments of the programme theory/theory of change or logical framework. Just under half of the terms of reference (not shown in the table) ask for assessments of unforeseen (positive or negative) results.

### Table 6: Additional questions in the terms of reference to effectiveness under the effectiveness criterion?

| Analysis of: | | |
|---|---|---|
| Specific break-downs of results | 63 | (79 %) |
| Mechanisms or explanations (regarding effectiveness) | 30 | (38 %) |
| The intervention's result or monitoring work | 18 | (23 %) |
| Programme theory, theory of change, logical framework | 17 | (21 %) |
| Can the objectives be reached in the future? | 6 | (8 %) |
| Have activities, outputs or results been delivered in time? | 4 | (5 %) |

Source: All 80 evaluations in the sample 2017–2019. Several choices are possible.

Part of the discussion in recent years in development assistance has been about the need for adaptivity in aid interventions, so it is interesting to see whether objectives and expected results are changed or adjusted during the period evaluated in the interventions. Table 7 shows that the objectives were adjusted or changed in just under half (44%) of the interventions evaluated in the sample.

[27] It is sometimes unclear whether the objectives that Sida wants to evaluate are those expected initially or whether Sida focuses on objectives or results that subsequently became of interest.

In 10 of these, they had been clarified or reconstructed in conjunction with the evaluation itself. Unclear statements of objectives rarely seem to have led to a reconstruction to facilitate evaluation.[28]

**Table 7: Have objectives/expected effects in the intervention been changed or adjusted in the period evaluated?**

| | | |
|---|---|---|
| No | 34 | (50 %) |
| Yes, at transitions between project periods/other reasons | 20 | (29 %) |
| Yes, reconstructed/clarified/restructured by evaluators | 10 | (15 %) |
| Not possible to see from the report/not stated | 4 | (6 %) |
| Total | 68 | (100 %) |

Source: All 68 decentralised evaluations in the sample 2017–2019 that focused on a single intervention. 12 evaluations focused on more than one intervention.

Table 8 gives the highest results stage assessed in the evaluations (cf. figure 1). Many of the evaluations (61 %) include an assessment of long-term objectives or impact. However, this is *not* the same as conducting a thorough analysis of impact. The section on impacts, which is included, is often short in relation to the other parts. The main emphasis of the evaluations is very much on the output and outcome stages, even in evaluations that have impact as the highest stage of their analysis. A third of the evaluations (36 %) give outcome as the highest stage of their analysis and therefore only assess effectiveness at the output and outcome levels.

[28] Sida say that they usually highlight the need for clarification of objectives with evaluators, and that this has been made clear in methods support and procurement.

**Table 8: Highest result stage in terms of time in the analysis of objectives/results (according to the evaluators)**

| | | |
|---|---|---|
| Impact | 49 | (61 %) |
| Outcome | 29 | (36 %) |
| Not possible to see | 2 | (3 %) |
| Total | 80 | (100 %) |

Source: All decentralised evaluations in the sample (2017, 2018, 2019).

However, there is no clear support in the terms of reference for the evaluations for not evaluating overall objectives (table 9), Even though the terms of reference are sometimes unclear, commissioners rarely state that they are *not* asking for such an assessment, saying instead that the interventions are to be assessed against the objectives set up in advance in the results framework or in line with what has been agreed.

**Table 9: Do the terms of reference explicitly exclude assessment of overall/long-term objectives?**

| | | |
|---|---|---|
| Yes | 4 | (7 %) |
| No | 47 | (80 %) |
| Unclear/text is contradictory/no terms of reference | 8 | (14 %) |
| Total | 59 | (100 %) |

Source: All 59 decentralised evaluations in the sample 2017–2019 that focused on a single intervention and where the conclusion on effectiveness could be identified in their report. Twelve evaluations assess more than one intervention. Nine evaluations have conclusions in which effectiveness cannot be seen.

## Summing up: objectives and remits in the evaluations

This survey demonstrates great pluralism regarding the overall goals in the interventions evaluated. The level of ambition differs widely between projects and the evaluators are often critical in part to how objectives/expected results have been formulated in terms of ambition, degree of specification, measurability, realism, consistency between documents and acceptance among the parties involved and time perspectives.[29] The overall objectives have sometimes been adjusted during the period of the intervention, which may be a sign of learning or adaptivity. The, according to the evaluators, unclear objectives rarely seen to have led to any clarification to make evaluation possible in accordance with the methods available for this (see Leeuw, 2002).

The terms of reference rarely state that they are *not* asking for an analysis of overarching goal attainment. The interventions are to be assessed against objectives set up in advance in results frameworks or in line with what has been agreed. The focus on unexpected results and impact is asked for in around half of the terms of reference. Just over a third ask for analyses of underlying explanations of goal attainment/expected results achieved, which is an important question if the purpose of the evaluations is learning.

[29] Far from all the terms of reference and evaluations talk explicitly and expressly about objectives and effectiveness in their evaluations. Instead they use concepts such as intended outcomes or, for example, outcome components. It is unclear what distinguishes objectives from intended outcomes here. One difference may perhaps be found in the more rules-anchored ideas and notions in the model for effectiveness evaluation about how objectives are formulated and arranged to be of use in management, monitoring and evaluation. The reference group for this study took the view that this is a matter of linguistic usage and that the difference is small in practice.

# 4. Conclusions about goal attainment

So what conclusions do the evaluators draw about the goal attainment of the interventions? The ambition in this chapter is to answer the study's question 1. The sample includes both decentralised and strategic evaluations.

## Two important limitations

Two limitations in the evaluations should be mentioned before presenting the conclusions about goal attainment. To some extent, they impair the conditions for a reproducible synthesis.

1. The assessments of goal attainment are often unclear, indirect, contradictory or mutlifaceted. Nine evaluations were so unclear that their conclusions could not be assessed.
2. Overall concerted assessments are not always made.

The reader's understanding is challenged when evaluation questions and conclusions are presented without any account of the intervention's initially stated objectives. The link between objectives and conclusions is unclear in several evaluations.

These factors give the review below a element of overall assessments that may, to some extent, impair the reproducibility of the synthesis concerning the effectiveness question (question 1).

As a further point, the assessments are rarely made in relation to the overarching goals, focusing instead on what has been achieved *so far* in each process considered, mainly at the output and outcome level.

The discussions and conclusions about effectiveness in the evaluations are exemplified in table 10. After that, the results of the review of the decentralised evaluations are presented, followed by the strategic evaluations.

### Table 10: Conclusions about effectiveness in four evaluations

| | |
|---|---|
| Wholly attained objectives/results | "The programme manages to achieve interesting, and sometimes even impressive results. […]Thesehighly relevant activities deliver direct results of high quality " (Gouzou et al, 2017). |
| Objectives/results mainly attained | "Although there were numerous delays, the bulk of the planned missions were completed, outputs were almost all produced, and the programme was largely effective. It was most effective in the area of … which has been completely internalised and institutionalised within….the same cannot be said of the …" (Moran et al, 2019). |
| Objectives/results mainly not attained | "The evaluators have attempted to assess results in relation to the five strategic objectives. Tentatively, this assessment indicates that [the intervention] has achieved specific objective number 2, and partly objective number 4. The specific objectives number 1 and 3 have not been reached. […]…the specific objective number 5…has made improvements… [but] the objective cannot be said to have been achieved." (Nyberg and Gunnarsson, 2019). |
| Objectives/results not attained | "There has been some progress made in achievement of outputs, but much less progress in the achievement of outcomes. While the difficult context in part has helps to explain the limited successes in the programme…[…] It has been difficult to find evidence of direct contributions that…[the intervention] has made to poverty reduction and to gender equality results" (Watkins and Azarbaijani-Moghaddam, 2019). |

Source: Four evaluations in the sample of decentralised evaluations 2017–2019.

## Decentralised evaluations

The compilation shows that the decentralised evaluations in the sample have generally reported good effectiveness and goal attainment.

Out of the 59 evaluations whose scale stage could be determined (table 11), as many as 50 state that expected objectives/results have been achieved, fully or in part, i.e. scale stages 1 or 2, up until the point in time when the evaluation was carried out. The most frequent conclusion (37 evaluations) is that, so far, the objectives have mainly been achieved (scale stage 2). Only nine interventions are not or mainly not effective (scale stages 3 or 4).

### Table 11: To what extent have objectives/expected results been achieved so far in the intervention, according to the evaluator?

| Scale stage | Specification | No | Share |
|---|---|---|---|
| 1. Wholly attained objectives/results | With the exception of one or a few single, marginal objectives/results, the assessment is that all objectives/results are attained or are on the way to being reached/will be reached, or: the intervention is described as being effective almost without any exceptions. | 13 | (16 %) |
| 2. Mainly attained objectives/results | The evaluator makes the assessment that the majority of objectives/expected results (or the objectives that the evaluator has stated as most important) are attained or are on the way to being reached/will be reached, or that the intervention is described as mainly or chiefly effective. | 37 | (46 %) |

| Scale stage | Specification | No | Share |
|---|---|---|---|
| 3. Mainly not attained objectives/results | The evaluator makes the assessment that the majority of objectives/ expected results, or the objectives that the evaluator has stated as most important, are not attained or on the way to being reached/will be reached, and/or: The intervention is described as mainly (but not wholly) ineffective. | 8 | (10 %) |
| 4. Not attained objectives/results | No or only a few objectives/results, or objectives/results of marginal importance, are judged to be attained/on the way to being reached/will be reached, and/or: The intervention is described as ineffective almost without any exception. | 1 | (1 %) |
| Not possible to see in the report/conclusions too unclear for determination of scale stage. | Information wanted cannot be seen in the report. Or The conclusions are very unclear. Or the evaluator has not made any assessment. | 9 | (11 %) |
| More than one evaluation object | Evaluates several different interventions and is therefore not assessed. | 12 | (15 %) |
| Total | | 80 | (100 %) |

Source: All 80 decentralised evaluations in the sample (2017, 2018, 2019).

Table 12 shows that the evaluators reported full effectiveness in a slightly larger share of project and organisation evaluations than programme evaluations.[30] At the same time, the observations are too few to enable any conclusions to be drawn about differences between different modalities regarding effectiveness.

### Table 12: Assessment of effectiveness per modality

| | Fulfilled? | | | | |
|---|---|---|---|---|---|
| Modality | Wholly | Mainly | Mainly not | Not | Number |
| Programmes | 1 (4 %) | 20 (80 %) | 4 (16 %) | 0 (0 %) | 25 (42 %) |
| Projects | 5 (33 %) | 10 (67 %) | 0 (0 %) | 0 (0 %) | 15 (25 %) |
| Organisation | 6 (50 %) | 5 (42 %) | 1 (8 %) | 0 (0 %) | 12 (20 %) |
| Strategy/plan | 1 (33 %) | 2 (67 %) | 0 (0 %) | 0 (0 %) | 3 (5 %) |
| Guarantee | 0 (0 %) | 0 (0 %) | 2 (100 %) | 0 (0 %) | 2 (3 %) |
| Unclear | 0 (0 %) | 1 (50 %) | 0 (0 %) | 1 (50 %) | 2 (3 %) |
| Total | | | | | 59 (100 %) |

Source: 59 decentralised evaluations in the sample. Evaluations of more than one intervention or with unclear conclusions have been removed.

Nor do I find any difference in reported effectiveness between the regions where Swedish development assistance is active (table 13). However, the share of interventions with fulfilled objectives in marginally higher in Eastern and Central Europe than in Sub-Saharan Africa.

[30] The two guarantee evaluations included are viewed as having displayed somewhat more feeble effectiveness. Programmes also tend to have a broader set of goals than projects.

**Table 13: Assessment of effectiveness per region**

| | Fulfilled? | | | | |
|---|---|---|---|---|---|
| Region | Wholly | Mainly | Mainly not | Not | Number |
| Sub-Saharan Africa | 4 (14 %) | 21 (72 %) | 4 (14 %) | 0 (0 %) | 29 (49 %) |
| Eastern and Central Europe | 3 (30 %) | 7 (70 %) | 0 (0 %) | 0 (0 %) | 10 (17 %) |
| Globally | 3 (33 %) | 4 (44 %) | 2 (22 %) | 0 (0 %) | 9 (15 %) |
| East/South and Central Asia | 1 (17 %) | 4 (67 %) | 0 (0 %) | 1 (17 %) | 6 (10 %) |
| MENA | 1 (33 %) | 0 (0 %) | 2 (67 %) | 0 (0 %) | 3 (5 %) |
| Latin America | 1 (50 %) | 1 (50 %) | 0 (0 %) | 0 (0 %) | 2 (3 %) |
| Total | | | | | 59 (100 %) |

Source: 59 decentralised evaluations in the sample. Evaluations of more than one intervention or with unclear conclusions have been removed. Here globally means more than one region.

Sida's evaluations are sometimes done in the form of "mid-term reviews" with the purpose "of assessing progress in an ongoing project to inform decisions on how project implementation may be adjusted and improved" (Sida's Evaluation Handbook). If a project is ongoing and at mid-term, perhaps there is not much that can be said about overall effectiveness?

In the sample 19 evaluations (24 %) are mid-term reviews.[31] They had a similar average budget to the other evaluations, so they cannot be seen as a simpler form of evaluation.[32]

[31] Data collected from Sida on 11 November 2020.
[32] The average budget for the 19 mid-term reviews in the sample was SEK 788 thousand, compared with around SEK 700 thousand for the whole sample.

In the median case the interventions evaluated had been running for four years. By type of evaluation, the median for the 19 mid-term evaluations is three years and five years for the other evaluations in the sample. The evaluators do not seem to assess effectiveness as higher or lower when the intervention has been running for a longer time (table 14). It therefore does not seem to be possible to obviously link effectiveness with whether the evaluation is a mid-term review or an intervention that has been running for a longer time.

**Table 14: Effectiveness and number of years evaluated**

| | Fulfilled? | | | | | |
|---|---|---|---|---|---|---|
| No years evaluated | Wholly | Mainly | Mainly not | Not | Number | MTR |
| 1–3 years | 8 (40 %) | 9 (45 %) | 3 (15 %) | 0 (0 %) | 20 | 10 |
| 4–6 years | 4 (13 %) | 22 (73 %) | 3 (10 %) | 1 (3 %) | 30 | 3 |
| 7–10 years | 1 (11 %) | 6 (67 %) | 2 (22 %) | 0 (0 %) | 9 | 0 |
| Total | 13 | 37 | 8 | 1 | 59 | 13 |

Source: 59 decentralised evaluations in the sample. Evaluations of more than one intervention and evaluations with unclear conclusions have been removed. MTR = mid-term review. Here the number of years evaluated corresponds to question 6 in the analytical framework.

One explanation of why effectiveness is not higher in interventions that have been running for a longer time could, for instance, be that the evaluations of interventions that have been running for a shorter time have assessed shorter-term results and that effectiveness can perhaps nevertheless be seen as increasing over time. This is supported by table 15, which shows that impact is evaluated to a slightly higher extent in evaluations of interventions that have been running for a longer time. However, the emphasis of these evaluations is on output and outcome (see the discussion in chapter 3 at table 8).

**Table 15: Number of years evaluated and analysis stages**

| Number of years evaluated | Output/ Outcome | Output/Outcome/ Impact | Number |
|---|---|---|---|
| 1–3 years | 12 (60 %) | 8 (40 %) | 20 |
| 4–6 years | 8 (27 %) | 22 (73 %) | 30 |
| 7–10 years | 2 (22 %) | 7 (78 %) | 9 |
| Total | 22 | 37 | 59 (100%) |

Source: 59 decentralised evaluations in the sample. Evaluations of more than one intervention and evaluations with unclear conclusions have been removed

The possibilities of drawing conclusions by making comparisons between more and less effective types of intervention on the basis of this material appear to be *limited*. As shown in table 11, only nine interventions are assessed as less effective. Table 16 lists these interventions.

**Table 16: Nine less effective interventions**

| Geography | Modality | Sector | Implemen-tation organisation | No of years evalu-ated |
|---|---|---|---|---|
| Zambia | Guarantee | Market, trade, entrepreneur-ship, innovation | Private | 5 |
| Uganda | Guarantee | Several areas | Private | 7 |
| Afghanistan | Unclear/ other | Gender equality | Multilateral/ bank, intergovern-mental | 4 |
| Global | Organisation | Conflict, peace, security | Civil society | 5 |

| Geography | Modality | Sector | Implementation organisation | No of years evaluated |
|---|---|---|---|---|
| Kenya | Programme | Sustainable urban development | Cross-border cooperation | 3 |
| Global | Programme | Market, business, commerce, innovation | Private | 5 |
| Regional, Sub-Saharan Africa | Programme | Water and sanitation | Public | 9 |
| Regional, MENA | Programme | Market, business, commerce, innovation | Multilateral/ bank, intergovernmental | 3 |
| Regional, MENA | Project | Market, business, commerce, innovation | Multilateral/ bank, intergovernmental | 2 |

Source: Nine decentralised evaluations in the sample where objectives/expected results have mainly not been achieved or have not been achieved at all.

Geographically table 16 shows that there have been less effective interventions on several continents, largely independent of how long the intervention has had time to take effect, in different sectors and among bilateral, regional and global interventions. The limited number of observations means that any differences may have arisen by chance, so no systematic differences can be traced. The clearest pattern in the sample of 80 decentralised evaluations of interventions in different contexts and through different modalities is that, according to the evaluators the interventions are almost always assessed as effective and *mainly* achieving their objectives.

## Strategic evaluations

All Sida's *strategic evaluations* in 2015–2019 that have assessed effectiveness (6) are also included in the compilation. Since these often have a broad focus on numerous interventions in different countries and present relatively complex conclusions about effectiveness, we have used a more rough stages while compiling the strategic evaluations.

The compilation of strategic evaluations demonstrates just about as strong reported effectiveness as in the decentralised evaluations.

### Table 17: Effectiveness of strategic evaluations (2015–2019)

| Evaluation | Overall assessment | Narrative conclusion: |
|---|---|---|
| Policy dialogue (2015) | Wholly or mainly fulfilled | "…policy dialogue has been an effective tool overall. Swedish embassies have been able to use it in diverse contexts to help achieve Sweden's development co-operation objectives.There is also clear evidence it has contributed to increased GE". |
| Capacity development (2015) | Wholly or mainly fulfilled | "…the Swedish support made a very important contribution to the development of the capacities of the partner organisations studied. […]On the whole the interventions are regarded as having medium to high overall effectiveness." |
| Guarantees (2016) | Wholly or mainly fulfilled (?) | ""…the guarantees of Sida are useful instruments that positively contribute to private sector development. […] In terms of effectiveness, impact and sustainability the intervention showed mixed results." |

| Evaluation | Overall assessment | Narrative conclusion: |
|---|---|---|
| ITP programmes (2017) | Wholly to mainly not fulfilled | "The ITP has been effective at capacity development of individuals but less effective at capacity development of organisations." |
| Challenge Funds (2018) | Wholly or mainly fulfilled | "…the intended outcomes for the majority of the funds have been broadly achieved." |
| Peace-building (2019) | Cannot be seen from the report/conclusions too unclear for scale stage determination | Conclusions cannot be summarised briefly in terms of a main conclusion. |

Source: Sample of six strategic evaluations 2015 to 2019.

The fact that he strategic evaluations analyse several interventions at the same time may have made it more difficult to drawn concerted and overall conclusions about effectiveness.

## Summing-up: conclusions about effectiveness

The evaluations of Sida's interventions (decentralised and strategic) almost always reported good goal attainment and effectiveness in interventions evaluated in the years studied. The most common conclusion in the decentralised evaluations is that, so far, objectives/expected results have mainly been fulfilled. It should, however, be borne in mind that the objectives achieved are generally at output and outcome level. It is rarer for the evaluators to assess that objectives at impact level are achieved, even though this seems to be asked for (see chapter 3).

The interventions described as less effective are so few that it is methodologically doubtful to draw conclusions about, for example, more or less effective project types.

# 5. The reliability of the evaluations in analysing effectiveness

This chapter is intended to answer the study's second question, about the reliability of the evaluations. My discussion starts from four criteria for reliable assessment of effectiveness:

1. Clearly defined objectives/expected results for the intervention (initially or reconstructed by the evaluator).
2. A fairly clear scope for the intervention, that has been running for so long that results can be captured.
3. A testable theory about how the intervention will achieve change (initially or reconstructed by the evaluator), possible external influencing factors in the context taken into account.
4. A well thought out evaluation design to assess whether it is the intervention that has caused or influenced observed results.

The starting point for the analysis is that a reliable assessment of effectiveness, with a primarily qualitative evaluation design, is made considerably more difficult if these conditions are not in place.[33]

Under each criterion, reasons are elaborated for why this particular criterion should be fulfilled to enable a reliable assessment of effectiveness to be made. In addition to these criteria, the evaluation should be transparent about how the question of effectiveness has been answered.

[33] We have assumed a qualitative evaluation method as all evaluations in the sample primarily use qualitative methods in their analyses of effectiveness. There are no, or almost no, regression analyses, econometric methods, experimental or quasi-experimental studies, but there is some use of descriptive statistics, monitoring of indicators and compilations of questionnaire data.

## Criterion 1: Clear, well-defined objectives

Objectives need to be specific and clear in evaluations so as to facilitate operationalisation and so as to be able to link them effectively to specific, observable, empirical evidence. Clear objectives make it easier to formulate good questions, to find a suitable evaluation design and useful metrics or indicators, to produce testable theories or hypotheses and to ask the right questions in interviews or focus groups. The conditions for good validity (measuring what is intended to be measured) then increase considerably. Clear objectives can also give a clearer scope and greater understanding between actors of what is being discussed (and why) in the evaluation process.

At the same time, it has long been known that objectives in public interventions, aid included, are often unclear (and formulated in a way that is not optimal for an evaluation), so they need to be reconstructed or clarified by evaluators as an early step in the evaluation work, preferably in close dialogue with the intervention or organisation being evaluated.

The evaluators in the material are, as indicated, often critical to how objectives/expected results have been formulated regarding matters like ambition, degree of specification, measurability, realism, consistency and time perspectives. However, the unclear objectives seem rarely to have led to any clarification work to strengthen evaluability in accordance with the methods available for doing so.

The conclusion is that this essential condition is often, but not always, absent at the time of the evaluation.

## Criterion 2: A clearly defined intervention that has run for a suitable period

Reliable assessments of effectiveness are easier to deliver in simple, cohesive interventions clearly defined geographically and in time. The activities should also have been running for such a long time that results would have had time to arise.[34]

In the sample, 16 evaluations (20 %) have a global perspective (defined as including at least two continents) and 18 had a regional focus (defined as including at least two countries on a continent being included in the project). Many of the regional organisations and programmes evaluated work in a very large number of countries. More than half the evaluations (46) have a national or local focus in the interventions studied. Another impression is that many interventions work in a relatively broad way as regards the number of areas and/or the number of objectives in the short- and medium-term (cf. Chapter 3). In almost 80 % (63) of the evaluations the commissioner asks for specific types of results or breakdowns of results/effectiveness. In 12 interventions (15 %) the evaluator is to assess effectiveness in more than one intervention when performing the evaluation.

The conclusion is that the analysis of effectiveness is complicated by the relatively large spread in terms of areas and geographically within the framework of the interventions and for each individual evaluation. Here there is a need to focus on effectiveness concerning important and priority results if it is to be possible to produce reliable assessments.

---

[34] It need not take a long time for aid interventions and other public programmes to generate results. Pawson (2013) argues that the dimension of time is one of a total of six dimensions creating complexity for evaluators. The dimension of time would not have been an example of complexity had it not been for the fact that a long-term perspective as such was required to assess results. Whether impact objectives can be expected to arise after 1 year or 7 years depends on the intervention and the context.

So, is it too early to assess overall effectiveness at the time when the evaluations are carried out? 56 interventions (70 %) had a total of *five years or more* of funding from Sida at the time of their evaluation. This does not mean that the period evaluated was that long. The average number of years evaluated was 4.5.[35] Nine interventions (11 %) had been running for 7 years or more, while eight interventions had been running for 2 years (cf. the discussion at table 14).

The answer to the question of whether the assessment was asked for too early differs between the interventions. The interventions have different levels of ambition. There are, however, no strong reasons not to evaluate against the overarching goals in projects that have been running for more than five years and that are, for example, intended to strengthen institutions' ways of working or capacity in civil society or in research conducted (common objectives in the sample). A relatively large share of the projects (especially the evaluations that are not mid-term reviews, with a median period of five years) could therefore have been assessed against overarching goals at the time when the evaluations were carried out. There are also a number of projects where this would probably have been too early (such as the mid-term reviews with a median duration of three years).

Based on length of intervention and level of ambition in the sample, a rough estimate is that it ought to have been possible to evaluate a third of the interventions thoroughly against the overarching goals at the time when they were evaluated. It can sometimes be appropriate to focus on results in the medium- or short-term, and in certain cases the outcome level can be the highest one. But there are no real reasons to never assess or to systematically refrain from assessing aid interventions in relation to the initially formulated overarching goals that were the starting point for funding. Nor does

[35] The interventions reviewed at mid term had been running for an average of 3.5 years (median 3 years). Six of them (32 %) have assessed results in some way at the impact level.

this mean that the focus must be on all the intervention's overarching goals. On the contrary, a systematic approach to evaluation sometimes requires a focus on one or a few central objectives.
The conclusion for criterion 2 is that the assessment of effectiveness is made more difficult by the lack of a clear geographical focus in the interventions, but that it ought to have been possible to evaluate a relatively large share of projects (at least 30 %) against the overarching goals at the time when the evaluations were carried out. Greater focus and adaptation of the objectives in terms of breadth and a greater focus on goals and objectives in the analyses appears to be necessary. Criterion 2 has only partly been fulfilled when the evaluations were carried out.

## Criterion 3: A theory of change

Criterion 3 concerns the relationship between the intervention and the expected result and is to form the basis for examining whether and, if so, how the intervention can lead to, leads to or influences the result envisaged. In the words of Pawson and Tilley (2001), *"Interventions are theories, and evaluation is the test."*

An absolute majority of the established design alternatives[36] for qualitative evaluation of effectiveness and impact now recommend that the work be constructed around a *theory of change.*[37] The purpose is to concentrate the collection of material around assumptions of change that are possible to test empirically (often in several stages) and that can make it possible to answer the "how and why" questions

[36] For example: Contribution Analysis (Mayne, 2012), Realist Evaluation (Pawson and Tilley, 1997), Process tracing (Beach and Pedersen, 2013), Systems based evaluation (Hummelbrunner, 2015), QUIP (Copestake et al, 2019). But also mixed method evaluation in which the quantitative component is experimental or quasi-experimental. A good brief introduction to the use of theory of change in evaluations is given in Vaessen et al (2020), page 157.
[37] Here I use the terms theory of change and programme theory synonymously.

that are important for learning.[38] A qualitative analysis of effectiveness not based on a theory of change risks being unfocused, having weaker validity, having difficulty linking intervention and results and not contributing to learning.[39] The theory of change must also weigh in factors that are external influences or drivers in the context.[40]

A distinction should, however, be made between the intervention's own "official" theory of change and the theory that the evaluator works with, develops and tests. As long ago as 1972, the evaluation researcher Carol Weiss described how these are often based on unclear and imprecise ideas about how to change reality when she introduced the discussion of the underlying assumptions of public programmes. This is a conclusion that has been verified many times. The fundamental assumption is bounded rationality (Lindblom, 1959, Stame, 2004, Weiss 2007). Public interventions are not formulated as wholly rational processes by evaluators and management experts, reflecting instead the political and complex settings where they were drafted. So a theory of change is not necessarily in place when the evaluator enters the scene; instead, just as is the case with objectives, the evaluator should be prepared to reconstruct or clarify the underlying theory, in several steps if

[38] Evaluators ask themselves, for instance, whether the theory is reasonable and in line with other knowledge in the area and whether the intervention has been implemented in line with the theory and can be verified empirically and what role is played by external influences. See rows 2–12 in table 19.

[39] Theories of change are also important as they enable common understanding between evaluators and intervention personnel regarding how or in what way the intervention is intended to lead to change. This is important in order to facilitate interaction, lay a foundation for communication about conclusions and reduce tensions in the work (Vaessen, 2020, page 157).

[40] The existence of a theory of change cannot be taken to show that the evaluation addresses causality. One recurring problem is that "theories of change" wholly or partly lack realistic and thought-out links to context and external influencing factors. The most central problem in theory of change evaluations is probably that, in practice, they often fail completely to present a theory of change, apart from some arrows and boxes (Ibid).

required, so as to enable a valid assessment to be made of effectiveness or results (see Leeuw, 2003, Weiss, 2007, Vedung, 2012).[41]

Nor is it a helpful conclusion in an evaluation that the theory of change for the intervention is unclear, unless it is followed by attempts to clarify (reconstruct) the theory of change and test it empirically.[42]

Table 18 shows that 16 of the evaluations in the sample probably had some kind of theory of change in place that was also tested empirically by the evaluator. In the light of, for example, Christopolos et al (2014) and previous literature presented above, we may assume that they were fairy imprecise.

**Table 18: Does the intervention have a theory of change that is tested against empirical data in the analysis of effectiveness?**

| | |
|---|---|
| Yes (probably) | 16 (27 %) |
| No (probably not) | 42 (71 %) |
| Unclear | 1 (2 %) |
| Total | 59 |

Source: 59 decentralised evaluations in the sample. Evaluations of more than one intervention or with unclear conclusions are not included. "Probably" is used since an aswer is often difficult to clearly extract from the reports.

The conclusion for criterion 3 is that there has often been some form of imprecise, sometimes reconstructed, theory of change, which has been discussed and assessed by the evaluators, but that is seldom tested properly against empirical data. Therefore the theory of

[41] There is, however, a limit beyond which it is no longer possible to reconstruct a theory of change that is too ill-conceived, absurd or unreasonable. But these are probably fairly unusual.

[42] Nevertheless, this conclusion is often drawn in Swedish aid evaluations. What should be the starting point in the initial phase of the work is then presented as a conclusion.

change is not integrated in the evaluator's analysis, probably making the analysis less focused, reducing validity and making it more difficult to link intervention and results.

## Criterion 4: An appropriate evaluation design

At a higher level the evaluator also needs an *evaluation design* for their analysis of effectiveness. That design brings together the evaluation question, theory of change, data and use of data in a unified whole according to an overarching logic.[43] The evaluator tries to answer the question of effectiveness by testing the sustainability of the theory of change against the data (documents, interview transcripts, questionnaire compilations or the like) according the starting points and processes specified in the design alternative with the requisite adaptation to the situation and context.

To be able to answer the question about the intervention's contributions in a reliable way, the design must take account both of the intervention and also of other potential influencing factors. Aid interventions generally function along with other influencing or supporting factors in a context.[44]

When I examined what design alternatives the evaluators refer to for the assessment of effectiveness, it turns out that half the evaluations (41, 51 %) do not contain a reference to any of the established design alternatives available for analysing effectiveness/expected results (table 19).

---

[43] The basic components of an evaluation design do not differ from those of research design in social science research, which consists, according to King, Keohane and Verba (1994), of a research question, theory, data and data use.

[44] The intervention is seldom alone in affecting or driving results. "Most development interventions are 'contributory causes'. They 'work' as part of a causal package in combination with other 'helping factors' such as stakeholder behaviour, related programmes and policies, institutional capacities, cultural factors or socio-economic trends. Designs and methods for IE need to be able to unpick these causal packages." (Stern et al 2012, page 7).

**Table 19: What design does the evaluator refer to in their analysis of effectiveness**

| | | |
|---|---|---|
| 1. | Other/Unclear/None | 41 (51 %) |
| 2. | Contribution analysis | 14 (18 %) |
| 3. | Combination of two alternatives (2–8 in the table) | 11 (14 %) |
| 4. | Theory based evaluation (general) | 7 (9 %) |
| 5. | Outcome harvesting | 4 (5 %) |
| 6. | Outcome mapping | 2 (3 %) |
| 7. | Most significant change | 1 (1 %) |
| 8. | Process tracing | 0 (0 %) |
| 9. | Realist evaluation | 0 (0 %) |
| 10. | Systems based Evaluation | 0 (0 %) |
| 11. | QCA (Qualitative Comparative Analysis) | 0 (0 %) |
| 12. | QUIP (The Qualitative Impact Assessment Protocol) | 0 (0 %) |
| 13. | Regression-based analysis (some form) | 0 (0 %) |
| 14. | Quasi-experiment (some form) | 0 (0 %) |
| 15. | Experimental design | 0 (0 %) |
| Total | | 80 |

Source: All 80 decentralised evaluations in the sample. Examples of overviews regarding design alternatives, methods and approaches for aid evaluation by quantitative and qualitative methods are Stern et al. (2012) and Vaessen et al. (2020).

Most of the evaluators who refer to a design alternative make partly unorthodox interpretations of what they refer to. John Maynes *Contribution Analysis* (Mayne, 2001) is the most common design alternative.[45]

Of the 14 evaluations stating Contribution Analysis, only four refer explicitly to John Maynes's articles, secondary literature or some other source giving information about how the method is used

[45] The fact that alternatives 13–15 are not referred to is expected, since all 80 evaluations are primarily qualitative. The evaluators often state that they use mixed methods. In practice, statistical methods and econometric analysis methods are almost never used; quantitative descriptive data and questionnaire compilations are found relatively often. Alternative 11 (QCA) is used comparatively, so it is only relevant when several interventions are compared at the same time.

specifically and only three mention Maynes's basic six-stage process.[46] Two (14 %) of the evaluations give brief reasons why they did not follow the process for the design they refer to. Do the evaluators specify in any other way how they have replied to the effectiveness question? Table 20 presents data for the 41 evaluations that did *not* refer to a design alternative.

**Table 20: Do the evaluators describe specifically and fully how the effectiveness question has been answered?**

| | |
|---|---|
| Yes | 8 (20 %) |
| No | 33 (77 %) |
| Total: | 41 |

Source: All decentralised evaluations in the sample that do not refer to any design in table 19.

20 % of the evaluations have described specifically and fully how the effectiveness question has been answered. In accordance with Sida's instructions they all have a method chapter and they often also have an inception report, which has to specify the consultant's method and implementation. Quite a lot of the evaluators have annexed an evaluation matrix.[47] Several evaluations state that there is one, but it has not been annexed to the report on publication. What is described there is often brief similar information between evaluation questions about how data has been gathered with weakly justified sources and indicators linked to each question.

The overall logic for the collection of material is often unclear. Descriptions or explanations of why specific methods and sources are used, how data contribute to answering particular questions or

[46] Contribution Analysis is relatively specifically defined in terms of six basic steps in the analysis (see, for example https://www.betterevaluation.org/en/plan/approach/contribution_analysis)

[47] These are a tool for planning and organising an evaluation. In practice this is often a table with a row for each evaluation question and columns dealing with method of data collection, source, indicators selected, etc. The idea is for the matrix to describe the operationalisation of each evaluation phase.

testing the intervention's theory of change and how external influencing factors are handled are incomplete. Often the collection of data is described, while the evaluation design is not. The collection of material often seems to be adapted to the evaluation commission as a whole, rather than to the individual questions.

70 % (56) of the evaluations take up or mention limitations of their method that directly affect the analysis of effectiveness.[48] Over and above the question of causality (see below), the most frequently discussed limitations are shortcomings in the interventions' data, reporting or documentation (30 evaluations), the results framework or logical framework (17) or the fact that the evaluators consider that the evaluation is being done too early for an analysis of effectiveness to be possible (9). So what is highlighted is the external factors that it has not been possible for them to influence themselves.

Conclusions about deficiencies in monitoring, reporting and documentation as an obstacle in evaluation work thus seem to have a long history in Swedish aid evaluation (see Peck, 2007, Forss et al 2008, Christoplos et al 2014). One would therefore assume that experienced evaluators of Sida's interventions see them as recurring givens that their own evaluation design solve or deal with rather than as unexpected obstacles. It may be a problem when interventions, as the evaluators seem to think, do not have any effective monitoring or baselines; but an evaluator ought to be able to adapt their design to the realities of development assistance and on the basis of learning from previous commissions. The evaluation of effectiveness does not need to be made so dependent on the intervention's monitoring and reporting.

So how do the evaluators handle the important question of causality (cf. Vedung, 2012)? A limited but partly illuminating picture can be given by charting the terminology in the reports. Three important terms in the evaluation vocabulary regarding assessments at the

[48] The evaluators take up several other limitations, but with an unclear or unstated link to the analysis of effectiveness.

outcome and impact levels are causality, counterfactual and additionality (see Sandahl and Petersson, 2016, or Sandberg and Faugert, 2012).[49] The terms can be used to verify that the evaluators are working in a particular way, to criticise or discuss alternative methodological approaches or to describe limitations in their own work. These terms occur to a limited extent in the sample. 70 evaluations make no mention of counterfactuality and additionality at all and 55 reports make no mention of the terms causal and causality. 26 of the 49 evaluations that claim to have analysed impacts never mention the terms causal and causality. A few of the reports talk about exogenous influence or exogenous factors.[50]

Only a third (25) of the evaluations make mention in their limitations section of any form of methodological limitations associated with causality and the analysis of effectiveness.[51] Among the reports that mention the question of causality, the most common observation is *that* causality has been difficult to deal with (in some cases this is claimed to be impossible) without going on to clarify *why* this has been difficult or *how* the authors have tried to deal with the matter and without discussing how this affects the reliability of the conclusions. No decentralised evaluation has a thorough discussion of whether it is possible to assess the effectiveness of the intervention at outcome and impact level without dealing with causality and external influencing factors.

---

[49] The concept of a counterfactual can also be of some relevance in qualitative evaluation since it opens up for a discussion about the alternative scenario without funding. A hypothetical exercise that is sometimes fruitful.

[50] Two terms that occur slightly more often are contribution and attribution. The terms have only been quantified for 2017 since the particular term contribution turned out not to reflect the discussion of causality in the reports. The term often crops up in contexts that are not associated with the question of causality or causal relations. In 2017 the term attribution occurred in 20 of 25 reports, and then not seldom in the form of a brief mention or disclaimer on the lines of "we focus on contribution, not attribution".

[51] Almost all reports have a sub-heading in their method chapter where the evaluator is expected to take up methodological limitations.

Of the 51 reports that had no discussions at all of methodological limitations concerning causality, 27 assessed impact. This means that they did not raise any methodological limitations regarding causality even though they focused partly on long-term results that are very likely to be affected by more factors than the intervention evaluated.

The overall review shows that 54 out of 80 evaluations do not describe how they assessed or tried to deal with causality and external potentially influencing factors in their analysis of effectiveness.

The conclusion is that, in general, the evaluators have not succeeded in convincing the reader that they have evaluated the *intervention's* effectiveness or contribution to effectiveness. Few evaluations in the sample have an evaluation design that is able to operationalise the question of effectiveness, test the theory of change and deal with causality in line with criterion 4.

## Summing up: the reliability of the evaluations

The conclusion from the review of the evaluations based on the criteria is that their assessments of effectiveness are not reliable.

The interventions have often not had clear and well-defined objectives, and the evaluations have rarely succeeded in reconstructing them. The analysis of effectiveness is made more difficult by a lack of geographical and area focus in evaluations' terms of reference and interventions. A majority of the evaluations do not refer to any established evaluation design, and in the other cases the description of the choice of design and the method are deficient. The impression is that the methodology is adapted to the commission as a whole rather than individual evaluation questions. The theory of change is seldom integrated with the analysis, which reduces validity and makes it more difficult to link intervention and results.

Few evaluations show how causality and external influencing factors have been handled. The evaluators have therefore not plausibly demonstrated that it is the intervention evaluated that has made a difference for, influenced or brought about the result. Consequently there is a risk that the reports describe changes in the contexts where the interventions operated that may be wholly or partly independent of the intervention evaluated.

# 6. Conclusions of previous studies

We have been able to note that the evaluations in the sample have mainly reported good effectiveness, but that the conclusions (as regards Sida's decentralised evaluations) are probably not reliable according to the four criteria for a reliable evaluation dealt with in the previous chapter. How surprising is that? Does it contradict previous studies, or is it what could be expected against the background of previous reviews of similar questions.

A review is presented below of some fifteen compilations, meta evaluations and analyses of Swedish aid evaluations from the past 25 years. The compilation focuses on 'grey literature' and does not claim to be exhaustive. It is based on simple but extensive internet and database searches to find as many items as possible. All compilations found have been included. The databases used are Openaid, Sida's publication database and the internet archive of the Swedish Development Forum (FuF). The publications found have been used to be able to locate new publications through their references to previous literature.

## Previous conclusions about effectiveness

The evaluation Promoting Development by Proxy (Riddell et al, 1995) analysed 37 projects in Sida's civil society support. One overall conclusion was that:

> "…the overwhelming majority either have achieved, or are well on the way to achieving, the stated and intermediate objectives for which Sida's NGO Division provided the funding."…: "project objectives provide little guidance to the overall development impact of the projects in question viewed more broadly. When the projects were judged against an increasing number of the nine broader criteria, their aggregate performance rating dopped progressively".

Nilén, Nilsson and Nybergs (2002) review of 171 Swedish CSO evaluations in 1999–2001 found that: *"[e]ven if these objectives are often only formulated in general terms, the evaluators generally conclude that the projects have reached their objective"*.

A meta evaluation of nine evaluations of Sida-funded networks of researchers in Africa drew the conclusion that:

> "What these nine evaluations have told us is that the projects are relevant and that they have achieved their immediate objectives. But they do not provide any substantial information on cost-effectiveness, impact or sustainability... Our conclusion is therefore that Sida's objective to support regional networks, in order to improve national research capacities, has not been dealt with and analysed sufficiently" (Carlsson and Wohlgemuth, 1996).

Three evaluation compilations carried out by Sida's Evaluation Unit (Peck, Jassey and Johansson, 2005; Johansson and Lindgren, 2006; Peck and Knippel, 2007) find high effectiveness in the majority of the evaluations.[52]

At country level Pain (2015) did a review of seven evaluations focusing on aid to Sweden's largest aid country Afghanistan. The interventions evaluated were judged to capture a large share of Sweden's portfolio to the country at that time. The study did not focus explicitly on effectiveness but drew the conclusion that:

---

[52] "In about half of the evaluations the level of effectiveness of the interventions was said to have been high." (34 evaluations, 2005), "For a majority of the evaluations, the level of effectiveness of the interventions…was said to have been high. In one-third of the cases the results were more mixed." (58 evaluations 2006), "Regarding the achievement of short- and medium-term objectives at the output and outcome level, effectiveness was found to be high for a majority of the interventions. However, the achievements of overarching objectives were more uncertain."(48 evaluations, 2007).

> ”… the interventions with a more instrumental focus have been more effective in achieving results. Those focussed more on creating understanding and reframing debates or building capacities and changing behaviours have achieved less…. while a number of interventions have improved access to public goods, such as education and health, and improved physical security, there have been few food security and income effects, limiting overall poverty reduction impacts.”

## Previous conclusions on the focus on overall objectives

One common conclusion in compilations of Swedish aid evaluations is that the studies have a weak focus on overall and central objectives in development assistance. The assessors seem not to have found a proper discussion in the evaluations about the broader contribution of the interventions for Swedish aid.[53]

A review of 80 Sida evaluations found that 30 tried, in some way, to demonstrate effects on poverty; but the evaluators *“were never asked to find out whether the prime objective of Swedish aid had been met or not”* (Eduards, 1995, quoted in Carlsson, 1998).

Another example is Tobisson and De Vylder (1997) who took the view that “scant attention is paid to poverty and gender issues in the country report documents and evaluations analysed. Somewhat surprisingly, given the long tradition of emphasis on poverty reduction in Swedish development cooperation…”

---

[53] This refers both to the link to a nationally produced strategy and the Government’s and Riksdag’s objectives for development assistance and focus on the overall vision in the interventions themselves.

A study of Sida's support to human rights and democracy, based on internal systems, documents and evaluations pointed to:

> "Insufficient attention to overall goals and how programmes are supposed to reach their impact. …(…)…Most projects appear well developed and considered with regard to their direct implementation, but less so when it comes to how they are supposed to impact on democracy and good governance in general." (Uggla and Wulfing, 2007).

A compilation of 84 Sida evaluations drew the conclusion that:

> "Less than half of the evaluation reports looked at the contribution of the interventions to poverty reduction, and most of those only analysed poverty reduction to a limited extent…(…)…A majority of interventions appear to lack an overall logic in their design and implementation. This often stems from a focus on activities. Many interventions (and evaluations) merely speculate on contribution to outcomes." (Christoplos, Liljelund Hedqvist and Rothman, 2014).

Another recurring point is the conclusion that Sida's interventions are often governed by unclear or imprecise formulations of objectives (see Bandstein, 2006 and Peck, 2007). Bandstein (2006) argued at the same time that assessments of effectiveness are part of what Sida personnel want to see most of from evaluation activities.

## Previous conclusions about reliability

Based on 177 evaluations, Forss (1994) was able to "refute" the hypothesis that (the majority of) Swedish aid evaluations make satisfactory assessments of effectiveness.

Another early study commissioned by Sida about the quality of its own evaluations found that a systematic analysis of evaluation quality gave a different picture from that obtained by asking programme managers directly about the reliability of evaluations. Programme managers made the assessment that Sida's evaluations were reliable, while the systematic analysis showed that the reports were extremely weak (Carlsson et al. 1997).

In the same year Carlsson and Forss (1997) analysed as many as 277 aid evaluations from Sida, the MFA and other Swedish aid agencies and found that the results were *"…discouraging. The evaluation system has over the years not delivered valid and reliable indications of the performance of development assistance"*.

A review of 34 of Sida's evaluation reports ten years later found that "[*m*]*ost of the evaluations cover effectiveness appropriately (62 %), although often in the sense of goal achievement at the output or near outcome stages"*. However, the authors took the view that the evaluations"*do not give the issue of attribution sufficient consideration, i.e. they do not show any empirical evidence of the intervention having an influence"*. The overall assessment of the evaluations was that "*evaluation quality assurance should be improved at Sida. There is a need for more and better empirical evidence and systematic use of such information in a majority of the reviewed reports. It is of particular concern that so few of the evaluations included enough information on the methods used. This made it difficult to assess whether the conclusions were reliable and clearly derived from the data*" (Forss et al, 2008).

A follow-up analysis of the quality of 114 decentralised evaluations five years later found that the*"…most prominent deficiency in this regard is the lack of applied methodologies for data collection and analysis. This is the basis for 'evidence based' evaluations and is critical when evaluating any level of evaluation or for distinguishing causality within a theory of change or from input to outputs and outcomes."* (La Guardia, 2013).

In a synthesis of 84 decentralised Sida evaluations Christopolos et al (2014) found that "*some programmes are far too ambitious in terms of geographic spread, complexity and number of intended outcomes to constitute a* *realistic or verifiable theory of change*."As regards the interventions' existing theories of change, a large part of the interventions have "*a weak linkage between activities, outcomes, results and impacts (the results chain) and do not have an overall logic… [...]…There is generally a failure to look critically at assumptions behind programming…[…] …It is often not clear whom the targeted direct and indirect beneficiaries and partners are, and how they are expected to benefit from the intervention*."

However the evaluations sometimes seem themselves *"to sometimes be used as an opportunity to effectively construct, reconstruct or otherwise draw attention to explicit or implicit theories of change"*.[54]

No systematic assessments of the quality of Sida's evaluations have been made after 2014. One exception is Burman (2017) who studied assessments of the sustainability criterion and found that the reliability of the assessment of the interventions was low.

## Summing-up: conclusions of previous studies

To sum up, it can be noted that, over time, Swedish aid evaluations commissioned by Sida have generally drawn the conclusion that effectiveness has been good in the short or medium term. Several previous analyses show that the evaluations do not focus on the long-term overall objectives of the interventions.

The fact that the interventions are also often considered to have unclear statements of their objectives may have been an obstacle to the possibility of evaluating effectiveness through clear operationalisation. External quality assessments often and time and again identify clear methodological deficiencies, including concerning questions of causality.

[54] I interpret this as saying that reconstructed theories of change were relatively uncommon when the study was done. One partial explanation may be that Sida had worked for a long time with Logical Framework Approach (LFA) as a management tool.

The above conclusions do nor apply solely to Swedish, or Sida-funded, aid. As regards individual aid interventions Riddell (2007) concluded that:

> "The available evidens suggests, quite strongly, that the clear majority of official aid projects achieve their immediate objectives. However, the evidence is far from comprehensive and likely to be biased in favour of a more flattering picture of the impact of project aid than is probably deserved."

According to Riddell (2007, page 186), there is evidence that donors' own internal assessments of the effectiveness of their projects overestimate the share of successful interventions.[55]

Colin, Sandström and Wennberg (2021) also give us a comparison with another policy area in their study of 110 evaluations of Swedish enterprise policy interventions in 2009–2019 and find that private consultants are "*predominantly more positive to the interventions evaluated.*"[56] The fact that *"the studies examined are not methodologically reliable should be set in relation to the fact that the evaluators have nonetheless been able to arrive at positive conclusions to so considerable an extent. What has affect the results is the specific fact that it is consultants that carry out the evaluations",* the authors assert.

A study of decentralised evaluations in Norwegian development assistance drew the conclusion that: *"decisions about Norwegian aid projects are being taken based on review findings and recommendations that are not always grounded on sound evidence".* (Chapman et al, 2017). According to the authors, the evaluations were used by decision-makers, but they were not reliable. Over time, previous analyses of different

---

[55] Donors' assessments based on internal monitoring should be kept apart from compilations of evaluations like this one. It is not clear from Riddell's compilation how he weighted donor reporting in relation to synthetic evaluation.
[56] Compared with evaluations carried out by the analysis agency the Swedish Agency for Growth Policy Analysis.

donors and large international organisations have also demonstrated general and widespread quality shortcomings in the aid area (Forss and Pollard, 2020).

To sum up, it can be stated that the conclusions of this report are clearly in line with previous studies.

# 7. Explanations of weak effectiveness

So, can we learn anything from the nine interventions where the evaluators reported weak effectiveness? Do the evaluations teach us anything about how projects should be implemented that aid actors, for instance a desk officer at an embassy, can benefit from in the further processing of an intervention or in future interventions? Can the evaluations explain why the interventions had weak effectiveness or make it understandable?

The focus here is on *hindering factors* highlighted by the evaluators. These are summarised in stylised form in table 21.

**Table 21: Impediments in weak effectiveness in nine interventions**

| Geography and sector | Impediments |
|---|---|
| Zambia, Market, business, commerce, innovation | Unclear management, donors not agreed on what the intervention would achieve, difficulties in changing behaviour in implementation stage, lack of resources and knowledge, Some financial risk aversion (on the target group side) and low competition. Weak monitoring |
| Afghanistan, Gender equality | Challenging context, weak management of the intervention, too ambitious objectives, lack of resources, weak ownership, not enough staff and high staff turnover, deficient risk analysis, monitoring and communication. |
| Global, Conflict, peace, security | Too ambitious objectives, few activities, lack of funding, capacity and focus. Weak model for following up results. |

| Geography and sector | Impediments |
|---|---|
| Kenya, Sustainable community development, urban development | Too ambitious projects in several sub-areas, delays, funding problems, inactivity on partner side, weak capacity interventions, changes of focus and direction. |
| Global, Market, business, commerce, innovation | Partly unclear direction and lack of focus in programme, differing interests and ownership in the target group, weak link to parallel initiatives and institutions. |
| Uganda, Several areas in same | Deficient analysis of basic problems, sector and demand. Partly incorrect finacial analysis. Absence of supplementary technical support and ownership in the implementation stage. |
| Regional Africa, Water and sanitation | Lack of focus, weak analyses and feasibility studies in several interventions included, weak coordination, wrong actors involved in sub-process, some interventions not financially sustainable, limited motivation among staff in one intervention, general problems concerning financial and sometimes material sustainability. |
| Regional MENA, Market, business, commerce, innovation | Deficient fundamental problem analysis, wrong actor participation, weak organisational capacity, deficient communication work, wrong activities, weak ownership and deficient internal effectiveness. |
| Regional MENA, Market, business, commerce, innovation | Low level of activity and delays, deficient project design in the form of a weak problem analysis resulting in wrong activities, deficient focus, too few and wrong staff, weak monitoring. |

Source: The nine decentralised evaluations in the sample where objectives/results have not been achieved either mainly or at all. The table describes hindering factors raised in various ways in the evaluations. It lists impediments without appraising their importance in relation to one another. The impediments do not always relate to the project as a whole and sometimes relate to specific interventions in the project.

When the above factors and categorised and grouped, we find a number of recurring conceivable partial explanations of deficient effectiveness in the interventions.

1. Deficient focus, management or leadership (6 interventions)
2. Organisational weakness or lack of capacity (6 interventions)
3. Weak ownership or demand (5 interventions)
4. Weak problem or fundamental analysis in the intervention (5 interventions)
5. Weak internal efficiency, low level of activity, delays (4 interventions)
6. Weak follow-up or monitoring (4 interventions)
7. Weak coordination, collaboration, wrong actor participation (3 interventions)
8. Weak communication (2 interventions)
9. Wrong activities (2 interventions)

The evaluators thus attach weight to a number of different mainly *intervention-centred factors* in combination, linked chiefly to the projects' fundamental analysis, management, organisation, implementation and monitoring. In terms of time, the impediments are in all of the *'policy cycle'* (Vedung, 2016) covering problem analysis, design and decisions on the intervention, the intervention and implementation, re-examination, monitoring, evaluation, dissemination of the re-examination, etc.

In explanatory evaluation a distinction is usually made between *theory errors* and *implementation errors* (Sandberg and Faugert, 2012). From that perspective the nine interventions seem to combine theory errors in the form of weak problem and fundamental analyses as well as weak ownership or demand with implementation errors such as deficiencies in management, focus and leadership, organisational weakness/lack of capacity and a low level of activity or weak internal effectiveness. In some cases, the interventions seem to have been wrongly planned and to then not have been able to be implemented according to plan on account of a combination of process-centred factors. The hindering factors centre in the implementation phase.

So do these evaluations teach the target group, desk officers at embassies for example, anything useful about what affects effectiveness and therefore can explain a lack of success?

Yes, partly; the evaluations supply several conceivable *partial explanations* of the weak effectiveness with very little focus on external explanations or mechanisms *in and in relation to* the intervention's context and setting. This is worth reflecting on since the discussion of aid constantly underlines the importance of context and context-specific mechanisms for understanding results and learning (two EBA reports that discuss them are Samoff, 2016 and Hydén, 2019).

Pawson and Tilley (1997) stress how contexts, mechanisms and results should be viewed as a whole in terms of time and space in qualitative theory-based evaluation.[57] There are seldom or never any exhaustive context-free explanations of results of aid interventions. To be able to explain or understand effectiveness and to bring about learning, conceivable external influencing factors in the context need to be taken into account and tested in a sufficiently thorough way with and in relation to process-centred factors. To find and "unpack" combined explanations (*contributory causes* or *causal packages*) some form of theory (see Stern et al, 2012) is needed that weighs in and tests alternative explanations and external factors (cf. chapter 6) as well as probably domain-specific knowledge. The theory of change as such should, as previously noted, often be reconstructed by the evaluator as they can rarely be taken direct from LFA matrices, for example.

---

[57] These are called CMO configurations i.e. Context–Mechanism–Outcome. The starting point (greatly simplified) is a hypothesis that programmes achieve outcomes (O) on account of some underlying mechanism (M), which is activated in some, but not all, contexts (C).

In part, the conclusions of an evaluation are a result of the analytical framework brought to the evaluation. The evaluators in the material often start from rational standard models of how interventions should be designed (the policy cycle, the SMART criteria, various logical frameworks) that are inserted in the analysis like a grid and bind the summative and formative analysis together. The risk is that the evaluators generate similar conclusions over time with constantly unclear problem analyses, unSMART objectives and unclear theories of change (the intervention does not live up to the standard model), but without generating new domain- or context-specific learning about how particular interventions function or do not function in their context.

These conclusions are not new. The Swedish National Audit Board (1988) found that there was some criticism of evaluation consultants in Sida and quotes one desk officer: *"the evaluations often present the same recipe dressed up in terms of evaluation jargon. Few evaluations cut through the surface. There are, however, exceptions"* (page 101).

A study of Sida's evaluations 10 years later (Carlsson et al. 1999) found: "*few examples of evaluations actually contributing something new in terms of knowledge."*

On the basis of a synthesis of 84 decentralised Sida evaluations Christopolos et al (2014) concluded that the "*Contexts and causes of poverty are not well analysed"*.

A nearby conclusion is that the hindering factors above (table 21) fairly seldom depend on sector- or domain-specific factors and mechanisms. The hindering factors in the nine evaluations concern the substance of the subject area to a small extent and seem, instead, to be based to a fairly high extent on assumptions about transferable management and evaluation models.

The evaluators take up a relatively large number of factors, but often without specifically concluding what factors specifically explain the weak effectiveness or make it understandable. In that sense the evaluators seldom make explanatory analyses, instead raising hindering (and to a smaller extent facilitating factors based on their own experience, statements in interview material or standard models. The contribution to learning is generically but insufficiently focused on general project methods that can, in practice, never make understandable or explain success or failure in an exhaustive way.

## Summing-up: explanations of weak effectiveness

The evaluations contribute possible partial explanations of limited effectiveness that can assist in some learning about the design, management, organisation, implementation and monitoring of the interventions. However, low weight is attached to external explanations or mechanisms in or in the vicinity of the broader context and setting of the intervention. Domain-specific factors, knowledge and mechanisms seem also to be missing to some extent in the analyses. The evaluators list a number of possible factors or impediments that can, in their view, improve the activities rather than analysing how or why results have not been achieved. In part, this limits the learning potential and usefulness of the evaluations.

# 8. Conclusions, recommendations and issues in the future

The main purpose of this study is to answer two overarching questions:

1. *What conclusions are drawn about effectiveness (or expected results achieved) in recently implemented evaluations of Swedish bilateral interventions?*
2. *Are the conclusions of the evaluations about effectiveness reliable and, if so, useful for learning?*

## Two conclusions

In chapter 4 the first question is answered as follows:

1. The sample of Sida's evaluations (decentralised and strategic) generally draws the conclusion the the interventions (mainly) reach their objectives, in so far as an assessment has been made.

This conclusion is in line with a large number of previous compilations of evaluations in Swedish and international aid (see chapter 6).

The conclusions of the evaluations I have reviewed are limited by the fact that the assessments are rarely made against overarching goals. The interventions described as less effective are few, and the variation is so small as to make it difficult in methodological terms to draw conclusions about more or less effective types of project or organisation, geographies or ways of working.

The potential for learning – new knowledge about what works where, when, how and why – is low when the evaluations' conclusions are similar. If almost all reports conclude that the interventions reach their objectives, there is simply not enough varying experiences to learn from. One may then question the justification behind a recurrent use of the effectiveness criterion.[58]

The second question is answered in chapter 5 by relating the sample to four criteria for reliable evaluation. The summary assessment is that:

2. The conclusions about effectiveness in the sample of Sida's decentralised evaluations are probably not reliable.

This conclusion is also in line with previous studies (chapter 6) that almost always question the reliability of Swedish and international evaluations of aid effectiveness. However, the question is whether this conclusion can be generalised to Sida's evaluations across the board and, if so, with what restrictions. Here it is important to take into account that this study does not treat other types of evaluation activities or other possible sources of information about effectiveness that need not be evaluations. It is also the case that the evaluation function (including quality assurance) is under continuous development, so previous evaluation samples do not necessarily reflect present quality. Further reflections about these questions are presented in chapter 9.

[58] A parallel is the evaluation criterion of *relevance* where the annual report for Sida's Evaluation Unit in 2019 drew the conclusion that, out of the 28 evaluations that studied the criterion, all had drawn the conclusion that the intervention had been relevant or highly relevant. This conclusion is also similar for the preceding years (2017 and 2018). It can be asked what a commissioning desk officer learns about relevance from a project evaluation, if they are not aware that almost all of Sida's interventions are assessed as relevant. What is the justification for using the criterion? See also Samoff (2021).

## Four recommendations

The analysis and conclusions in previous chapters lead to four recommendations. They are addressed in the first place to Sida's Evaluation Unit, but are also addressed to Sida's leadership and the MFA representing internal and external demand for more reliable evaluations against overarching goals.

**Recommendation 1:** The methodological quality of decentralised evaluations should be strengthened considerably (see chapter 5).

A future improvement of methodological quality is important to make it possible to rely on the conclusions of Sida's evaluations at all levels in development assistance. Without reliable decentralised evaluations, knowledge is missing about the effectiveness of Swedish aid at intervention level. It is not likely that partner evaluations, the EBA's evaluations, research or internal monitoring work can compensate for this.

**Recommendation 2:** More interventions should be evaluated against overarchng and long-term objectives (see chapters 3 and 4).

Without evaluations against overarching expected objectives, no full assessment of effectiveness is made. For the future it is important to make clear why Sida evaluates its interventions against overarching goals to such a small extent. In this study I have shown that this is probably not a matter of a lack of demand from Sida's desk officers and other commissioners.

An alternative, additional to analyses of objectives at the overarching level, can be to analyse the intervention within a results chain analysis (Sandahl, 2019). This makes the evaluation less dependent on timing. Sometimes, a focus on results in the short or medium term is appropriate whereas in some cases, outcome may be the highest level. But it is problematic to systematically refrain from assessing aid interventions against initially formulated overarching goals. My conclusion is that around a third of the interventions in the sample could have been assessed against overarching goals.

Today Swedish aid focuses to a fairly high degree on core support or regional and global forms of support that are difficult to evaluate. Sida should therefore assess the evaluability of these interventions with respect to the possibilities of evaluating and assessing goal attainment and effectiveness of Swedish aid in the future (cf. the discussion in chapter 5).

The fact that evaluations are a support for intervention management; are designed for special needs; are to be participatory and are sometimes conducted at the mid-point of the project cycle does not constitute strong arguments for not evaluating against overall objectives. There is no essential contradiction between Sida's operations-centred learning focus and the possibility of producing full assessments of goal attainment.

**Recommendation 3:** Assessments of expected results should be supplemented to a greater extent by analyses of unexpected results and potential side effects (see chapter 3).

It has long been established that public interventions and projects can have negative unforeseen side effects, and this also applies to aid. There are also strong ethical reasons for assessing unforeseen effects of aid. In part, these analyses can be conducted as separate studies so as not to make Sida's decentralised evaluations more unfocused.

**Recommendation 4** In the future, the evaluations should focus around goal attainment and on trying to explain and understand how and why interventions do (not) reach their objectives (see chapters 2, 3 and 7 and annex 4).

The learning potential from the evaluations can be strengthened by focusing to a higher degree specifically on goal attainment/expected results and on questions concerning why and how results are reached or not reached; what factors or mechanisms explain success and failure; and questions concerning who benefited from the intervention, where, when, how and why.

Preferably starting from theory-based perspective (Pawson and Tilley, 1997, Stame, 2004, Mayne, 2012, Stern, 2012, Beach and Brun Pedersen, 2013), the evaluations should focus on learning by, as far as possible, trying to open the "black box" of mechanisms, context and processes, which can contribute explanations of the probably differing results of aid in different contexts, at different times and for different target groups.

## Five questions for further studies

There are a large number of nearby and related questions about Swedish aid, evaluation, monitoring, reporting and effectiveness that have arisen in the course of the work but that have not been accommodated in this study. The following five questions should be in particular focus for future analysis.

1. How is goal attainment assessed more generally in Sweden's bilateral development cooperation? What supporting information do Sida, the Folke Bernadotte Academy (FBA), Swedfund and others use, over and above evaluations, to assess and report goal attainment to the Riksdag and the Government? How reliable is this supporting information and these assessments?
2. How can and should Sida's information about results (based on evaluation, monitoring and reporting) be aggregated at strategy and country level?
3. How can Sida and other donors assess unforeseen results and side effects in a more comprehensive, systematic and reliable way? How are they evaluated today? How can and should this be done?
4. How reliable are the conclusions of Sida's strategic and partner-led evaluations?

5. How can the challenging assessments of causality in evaluations be developed to also achieve a ‘good enough’ level in the evaluation of small or limited aid interventions, or when the evaluation budget is limited?

# 9. Concluding reflections about reliability

One important message of this report is that Sida should increase its ambitions to conduct reliable evaluations of the effectiveness of interventions, and to attach more weight to overarching goals in doing so. Is this a reasonable and realistic expectation? In this concluding chapter a number of reflections are presented about possible causes of deficient reliability in four areas. The causes are discussed in an open-minded way and without any claim to put forward final or definitive conclusions. They should be viewed as tentative starting points or initial positions for further discussion and analysis.

There is, today, a societal and political expectation that Sida will evaluate its interventions. It is probably unrealistic to expect the Government or the Riksdag to tone down their demands for evaluation in this area.

One of the most used and established definitions of evaluation, also used by Sida, is: *"a systematic and objective approach of determining the merit, worth or value of an object"*. The idea of evaluation thus contains an expectation of a systematic approach, that evaluation is conducted in a thorough, independent and (fairly) exhaustive way so that the value of the intervention can be established as far as possible. On its website Sida describes its own evaluation work an "important tool": *"Evaluation differs from monitoring activities where evaluation is occasional in timing, focus on longer term outcomes and impact and has the possibility of analysing and valuing the results from a variety of information sources."*[59]

This study has not examined *why* the evaluations' assessments of effectiveness are not reliable or whether there is a link between conclusions and methodology. Nor have I studied the reliability of

[59] https://www.sida.se/en/for-partners/partnership-with-sida/evaluation-of-swedish-development-cooperation

the strategic or partner-led evaluations that Sida views as important in work on evaluating Swedish development assistance. However, well-developed evaluation work requires further analysis of what the deficient reliability of the decentralised evaluations is due to. The continuity of the problems underlines how important it is that those responsible take a comprehensive approach to the question.

The limitations are potentially both on the demand side (within Sida) and on the supply side (among evaluation consultants and in the market for evaluation services), so the answer should be sought in both these spheres and in their interaction. Neither the commissioners nor the evaluators are assumed to have full control of the end product even though the *person responsible* for quality in a particular evaluation can always be considered to be the main author of the report.

Four areas that can be possible starting points in an explanatory analysis of Sida's decentralised evaluations concern:[60]

1. Demand, needs and incentives.
2. The system for procurement, call-off orders, feedback and following up of evaluations.
3. Supply in the evaluation market.
4. Knowledge about evaluation and about learning and development in activities.

---

[60] Every evaluation is based on demand that takes shape ahead of the procurement process or in a call-off order based on institutionalised ways of working, knowledge and experience within the framework of existing systems and structures. The call-of order generates a "reply" from the supply side in the form of a report that is received, followed up, given a reply and published and forms the basis for learning in the form of new studies, future framework agreements for evaluations, etc. In theory the system can have components of all of rational efficiency maximisation, institutional inertia, being track-bound, ”muddling through" and decision-making using a "logic of suitability" (Sundström, 2001).

These areas focus on the commissioning level and its relationship with evaluators in the market for evaluation services. The starting point is that the reliability of Sida's evaluations is most likely to be affected by a number of interacting factors, but that some aspects are more important than others. This means that the problem is not solved by taking single measures and that it will be necessary, following a thorough analysis, to consider a combination of several interacting measures. Swedish aid evaluation is also part of internationally standardised activities in which OECD/DAC is a central actor and the market for evaluation services is international, so account should also be taken of the international level.

## 1. Demand, needs, incentives

Sida applies a utilisation-focused evaluation approach. The focus is on identifying the intended users of a specific evaluation, on there being a clear demand and a specific need. At an overarching level, reliability requires that the commissioner has clear expectations about reliability and gains acceptance for their expectations in the contract and dialogue. The quality of the end product is very much an outcome of the quality demanded and ordered by the client. The factors that affect this include the commissioner's prior knowledge and the clarity and level of detail of the commission. However, it also requires a continual process working for reliability that responds to and intervenes against deficiencies.[61]

In evaluation contexts it can be assumed that commissioners' focus regarding evaluation criteria and issues can affect quality and reliability. Quality is dependent on the commissioner's ability to balance an often pluralistic set of expectations against the need for

[61] Purchases of evaluations can be analysed in the same way as other markets (Nielsen et al, 2019) and start from the purchaser's influence and position in the market for evaluation services. It is then influenced by the number of other purchasers, the financial size of purchases, differences in supply between competitors, price sensitivity, the opportunities for changing supplier or replacing one service with another.

focus and clarity that makes methodological stringency possible.[62] OECD/DAC has an influence on the demand side through the evaluation criteria that Sida and others work with. Because even though both the OECD and Sida state that the criteria should be used with judgment it is likely that the application of the criteria has driven or at least maintained a lack of focus in Sida's decentralised evaluations. This may, in turn, have limited the possibilities of making thorough analyses of evaluation questions such as effectiveness (see Burman and Hårsmar, 2015).[63]

The fact that the evaluations are often commissioned and carried out during the period of the intervention or in conjunction with the ending of the intervention (cf. Chapter 3) often makes an evaluation against overarching goals more difficult. A relatively large number of Sida's terms of reference specify how the interventions are to be carried out methodologically, who is to collect what data when and similar matters. This can make it more complicated for evaluators to do a good job.[64]

Are there weaknesses in Sida's guidelines or policy that can explain the deficiencies? Sida's Evaluation Handbook (2020), which has evaluators and commissioners as target groups, clearly states the importance of reliability and integrity in the evaluator and refers to OECD/DAC's quality documents that stress the importance of answering questions on the basis of reliable evidence. Even though the Handbook hardly discusses the effectiveness criterion, it does

---

[62] At the same time, it can be asserted that one important skill possessed by an experienced evaluator is responding to and communicating the difficulty of, for example, many evaluation questions given a particular budget and time frame.
[63] The OECD cannot be made responsible, since a government agency should, when applying internationally agreed, non-binding, criteria, ensure that they are implemented as a well-considered "translation" to Swedish conditions. Commissioners in the organisation need to understand how the criteria are to be used and how best to set priorities between criteria.
[64] For example several terms of reference in the sample lay down that the evaluators have to conduct a questionnaire survey, at the same time as the commissioner does not make clear why the survey is needed in that specific case.

underline methodological questions and reliability. This probably shows that better policy and governing documents have a fairly limited role in further development work.

At the same time, assessments of effectiveness in specific organisational contexts can be viewed in relation to other organisational objectives, internal power relations and incentives. It is not always obvious that reliability trumps the "right" conclusion, just as depth does not always trump breadth. In a study of the set of incentives at Sida, Nobel laureate Elinor Ostrom took the view that *"No effective demand exists for meaningful evaluations of development assistance".* (Ostrom et al, 2002, page 247).[65] Ostrom's assertion that the aggregate incentives among influential actors lead to "watered-down" evaluation processes where the actors whose voices are heard (donors, evaluation consultants and staff in the interventions) may lack incentives to describe the intervention in a critical and reliable way, to suggest termination of an intervention, etc.

Several researchers argue that evaluation has its own political economy.

> "Evaluation becomes an activity that takes place in a political context. The policies and programmes that are being evaluated are proposed, defined, debated, enacted and funded through political processes. The evaluation itself has a political stance because it makes implicit political statements about issues such as such as legitimacy of programme goals and of project strategies, and the usefulness of various implementation strategies. […] An evaluation of the effectiveness of programmes should perhaps also incorporate scores for effectiveness in the politics of organisational survival." (Carlsson et al, 1994, page 176).

[65] It is unclear whether Ostrom drew her conclusions solely based on studies of Sida.

The commissioners place commissions with evaluators who act within a primarily commercial logic and quality assurance rather than a scientific logic and quality assurance, where the incentives to provide critical comments to a loyal customer can be discussed.[66]

On the commissioning side, the view may sometimes be formed that rough assessments in the short- or medium-term of the results are sufficient for the needs commissioners have; perhaps desk officers do not always show an interest in methodological questions, etc.[67]

*If* the conclusion is that it is factors on the demand side, incentives or political economy that have a negative impact on reliability, the solutions can involve steps like:

- Action to strengthen Sida as a customer in the market for evaluations (Nielsen et al 2019).
- Agreements between the MFA and Sida and between Sida and partners to the effect that decision-making, where based on evaluations, has to be founded on good evidence.
- Measures to convince managers and desk officers of the importance of methodology in evaluations.
- Further work on focusing and defining the scope of evaluations and the terms of reference's number of evaluation questions linked to their complexity, the budget and implementation time.
- Development of processes to set priorities between the OECD evaluation criteria and issues in specific evaluations (Burman and Hårsmar, 2015).
- Measures to restrict any internal political influence on the design or conclusions of evaluations.

---

[66] Sida states that it is partners and not Sida that are evaluated and that this weakens that type of incentive. At the same time Sida should be viewed as a stakeholder even when evaluating partners.
[67] This conclusion is in line with Carlsson et al. 1997 who reported clear differences between an external quality assessment of Sida's evaluations and the assessment made by the desk officers themselves.

- Further codification of requirements for integrity, objectivity and independence.
- Reinforced processes for clarifying and defining the scope of interventions' objectives before evaluation work is started.

## 2. The system for procurement, feedback and follow-up

Difficulties in ensuring reliability can be associated with systems and processes for procurement/call-off orders (framework agreements), the budgets of the evaluations, the evaluator's ownership, feedback to suppliers and resourcing.

It is probably important how Sida's *framework agreement* for evaluation consultants are designed in terms of incentives to deliver reliable evaluations and that Sida is given access via the framework agreement to the best possible evaluators internationally. One basic question is the extent to which not delivering good quality has repercussions, given that reliability problems have recurred over time. The framework agreement for evaluation services should also balance the need for exposure to competition against the importance of it being easy for staff to commission evaluations.[68]

When Sida had an evaluation done of its framework agreement for evaluations in 2014, the conclusion was that the framework agreement had not *"been a sufficient mechanism in and of itself to improve the quality of evaluations"* (LaGuardia et al, 2014, page 9). More specifically, the ranking system for suppliers had not been an *"adequate procurement model for promoting quality evaluation services"* (page 10). The framework agreement is different today, and we do not have supporting information to assess whether it promotes or

[68] Repeated exposure to competition can make more work for the commissioner, which may lead to fewer evaluations being conducted, at the same time as a lack of competition under the framework agreement may reduce quality.

counters quality. Undoubtedly, procurement under a framework agreement and the design of the agreement have great potential to exercise an influence for the better or the worse. The framework agreement and its follow-up is also a tool where Sida's Evaluation Unit has great possibilities of influencing the quality of decentralised evaluations.

One important question is whether the *evaluations' budgets* are sufficiently well dimensioned (especially per evaluation question). The average price for Sida's decentralised evaluations has increased from around SEK 500 thousand in 2011–2014 (La Gaurdia et al, 2014) to around SEK 700 thousand in 2017–2019 (median SEK 680 thousand, minimum SEK 230 thousand, maximum SEK 1.5m). As an estimate, the evaluation team therefore have around 15 working weeks, or just under 4 months, to carry out an average evaluation without having to reduce their hourly rate.[69] A reliable evaluation needs to have enough time for the different phases of their work: planning, collecting data, analysis, reporting and quality assurance.

Forss et al (2006) studied Sida's evaluations but found no clear correlation between cost and quality. The best evaluations were not always the most expensive ones, and it was not considered that the quality problems could be addressed with a bigger budget. If Sida is not able to reduce the number of questions in decentralised evaluations, it should perhaps analyse whether the budget per question or evaluation criterion needs to be further increased.

A correctly considered budget is not all that is needed to be able to achieve an evaluation of good quality; this also requires calendar time and control, ownership, by the evaluators of their own use of time. It is not obvious that these go hand in hand. If the embassy specifies, in advance, one week of field work for the evaluation team without this being rooted in the design and use of time needed by the

[69] We have made our calculations using hourly rate of SEK 1,200 per hour for consultant fees but have not included costs in the project apart from staff. With three people in the team, that makes around 5 working weeks per person.

evaluator, this reduces the possibilities of doing a good job, even though the budget as a whole is correctly considered. If this is a problem, Sida can, for example, consider increasing the average calendar time available for decentralised evaluations, or the Evaluation Unit can, in its advice, stress the evaluator's ownership of their own use of time in terms of data collection in the field, for instance.

Even if central support is available via Sida's Evaluation Unit in forms including the handbook, guides, templates and the possibility of obtaining advice, the conditions for reliability are affected by the fact that commissioners at embassies or among desk officers are often inexperienced commissioners of evaluations (cf. Burman and Hårsmar, 2015 and Burman, 2017).

After an inception report and draft report are delivered, it is important that the evaluator is given a thorough *response* and comments of a high standard. However, the whole situation contains an imbalance since the evaluator generally has long experience of evaluation and the context, while the desk officer seldom does.

I have understood that Sida do not consider that they have sufficient resources to support or quality review all decentralised evaluations (around 30 per year) extensively. Previous analyses of Sida's evaluations (Forss et al, 2007 and La Guardia et al, 2014) reported limited capacity among Sida's officers to supply technical support in the evaluation process. At the same time, the commissioner of an evaluation ought to have the ability and capacity to assess what is a good or bad delivered service. If the evaluator seldom or never receives feedback on their delivery or if there is no link between quality and the content of the feedback, this is likely to undermine reliability in the long term. The value of asking for reliability decreases if, in practice, the purchaser does not respond to quality shortcomings. But nor do I have any updated knowledge about what the situation is like today.

However an analysis should be made of whether the decentralised evaluations do not require a strong component of central responsibility if there is a wish to improve the reliability of the decentralised evaluations. Then an analysis is needed of whether there are sufficient *financial and staff resources* at Sida's Evaluation Unit to ensure the quality of the decentralised evaluations. The situation has changed over time, but when OECD/DAC conducted a survey of the evaluation functions of a large number of bilateral and multilateral donors in 2016, Sida's Evaluation Unit reported the lowest budget of all OECD/DAC donors as a share of the total aid budget (OECD/DAC, 2016 page 230).[70]

*If* the conclusion arrived at is that it is the framework agreement, the procurement process, feedback, resources and the follow-up of the evaluations that are holding reliability back, the measures can include:

- Review of the framework agreement focusing on trying to strengthen the incentives to deliver reliable evaluation reports.
- A stronger process for responses and feedback with evaluation expertise on reports delivered. For example by strengthening Sida's Evaluation Unit's capacity and possibilities of supporting commissioners.
- Increased financial and staff resources to Sida's Evaluation Unit (to support work on ensuring quality in decentralised evaluations).
- Further strengthening the possibilities for evaluators to plan their own time and design their own scheme and implementation within the framework of their evaluation commission, including the collection of material in the field, and to do so on the basis of their own expertise in the subject area.

---

[70] One reason for this may be that Sida's evaluation budgets are taken from the policy appropriation. Since then Sida's Evaluation Unit has been reorganised, so the resource situation may have changed.

- Increased calendar time as a standard for every evaluation.
- Increased evaluation budget per question or evaluation criterion.

## 3. Supply of evaluation services

Successful procurement of evaluation services is also influenced by the seller's position in the market in relation to the purchaser and the existing competition (Nielsen et al 2019).[71] To be more specific, the supply of evaluators in Sida's decentralised evaluations is governed by who is awarded a framework agreement with Sida following a framework agreement procurement and the particular evaluators supplied under the framework agreement. The quality review of Sida's evaluations in 2014 (La Guardia et al) reported an unchanged situation over the period 2008–2013 concerning:

> "A limited number of qualified consultants, and missing skills and capacities among the evaluators with which Sida works, which might reflect a lack of competition and little variety among consultants." (Ibid, page 96).

I have no data about what the situation is like today regarding, for example, the turnover of contracted consultants over time or the development of hourly rates for evaluation services.[72]

One indication that the competition may *potentially* be large is that aid evaluation is an international activity that does not need to be carried out by Swedes or via Swedish companies. The service is not tied to a specific geography, except when collecting data. Here an analysis should perhaps be made of whether there are "thresholds" in the

[71] The number of sellers, their relative size and capacity, differences in uniqueness and quality between actors, threats from new actors, thresholds in the market, etc.

[72] The hourly rates could have given an indication of the level of competition and the market's degree of maturity.

framework agreement system that mean that Sida is not able to benefit fully from international competition and capacity in the area of evaluation. An analysis can be made of whether there are ways, on the part of consultancies, to recruit for framework agreements that reduce the inflow of international methodological and evaluation expertise or that limit the turnover or inflow of evaluation expertise. In this context an assessment should be made of how much weight to attach to Sida-specific evaluation experience since such requirements may reward suppliers who have not previously delivered good quality.

*If* the conclusion arrived at is that deficiencies in the supply of evaluators or evaluation firms may have affected reliability, this can lead to measures such as:

- A change of process for future framework agreement procurement focusing on broadening the supply of evaluators and increasing competition with a focus on stronger evaluation and methodological knowledge.
- Reinforced action to reach out to new groups of suppliers of evaluation internationally in conjunction with framework agreement procurement.
- Continued and expanded dialogue with existing framework agreement actors about broadening the supply of evaluators focusing on strengthening and broadening methodological expertise among both evaluators and quality examiners.

# 4. Knowledge, learning and development in operations

The three areas discussed so far are very much to do with the *forms* for Sida's purchases of evaluations. However, the content of the work is influenced to a very high degree by the knowledge in the system and the learning that takes place in the course of operations.

Assessing goal attainment in a public intervention is a high-level task that requires knowledge of advanced evaluation methodology, subject area issues and context. At the same time, there is, at present, no accreditation system for evaluators; there are no formal requirements for anyone who wants to call themselves an evaluator and the quality assurance systems in existence (in OECD DAC, for instance) cannot and should not specify how to conduct a specific evaluation.[73]

One important quality function in the decentralised evaluation is the quality assurance functions of framework agreement suppliers. At the same time, there are several potential pitfalls if the person responsible for quality comes into the work too late, if the quality audit does not have a systematic approach or if the person responsible for quality does not use accepted criteria or adapts to an informally expected (lower) level of quality than they are used to working at. Evaluators may lack incentives to listen to the person responsible for quality and the quality review may be formalistic or lack a substantive or sector-specific perspective. Alternatively, a quality audit may be "window dressing" with a limited impact on the actual evaluation work.

One overall impression from this report's review of 80 evaluations is that contemporary thinking about evaluation of aid interventions as regards approaches, design choice and methodology does not really make an impact in Sida's decentralised evaluations. Some aspects of this are that the reports seldom propose any real evaluation design and the lack of references to modern evaluation methodology and literature (which have often been produced for qualitative analyses in the area of development assistance, see Vaessen et al, 2020 or Stern et al, 2012, for example). But also the

[73] The discussion of training, professionalisation and accreditation of evaluation and evaluators has been going on for almost as long as the discussion about evaluation (see Schwandt, 2019). An evaluation must be adapted to the particular intervention and the commission, which limits the usefulness of generic check-lists and centrally defined processes.

lack of transparency about how particular evaluation questions have been answered; discussions about causal connections are, when present, brief and deficient. Many of the design alternatives now available for the qualitative or quantitative evaluation of aid interventions were not used in the years we studied. In the cases where the report specifies the use of an established methodology, the upshot is often that it is not applied in practice.

As regards Sida's learning about evaluation activities, it can be noted that this study is one of a series of studies that have drawn similar conclusions about the quality of Sida's evaluations over the past 25 years (Burman, 2017, Christoplos, 2014, LaGuardia et al, 2014, Forss et al, 2008, Carlsson et al, 1997, Forss, 1994). The impression is that these studies have not had any great effect on the quality of the evaluations in answering the question of effectiveness. This is despite the fact the majority of the studies had been commissioned by, and were therefore wanted in some way, by Sida itself. This does not mean that Sida has not learned from the studies, but it does mean that this learning has not resulted in clearly more reliable evaluations.

At the same time, it is probably challenging to bring about learning out in Sida's organisation, so responsibility needs to be taken at central level for decentralised evaluations. An individual desk officer at Sida quite seldom commissions evaluations, and the lessons learned may perhaps only be used on a few occasions during a career at the agency.[74]

As regards the evaluators' learning, it can be noted that there appears to have been a tendency to raise similar obstacles to the implementation of evaluations, focusing on unclear objectives,

---

[74] Around 4 % of Sida's employees (25 of 643 full-time equivalents 2020) can be expected to be responsible for commissioning a decentralised evaluation in a particular year. Assuming that the need is spread and varied across the organisation and over the years and that few staff members work for more than ten years at the agency, it is likely that few employees (outside the Evaluation Unit) will have time to order more than two to three evaluations during their career at the agency and that they will probably do so at fairly long intervals.

deficient results frameworks, the absence of baselines, unclear theories of change and weak internal monitoring in projects. When it is, to some extent, the same evaluators who have carried out these evaluations over time and drawn the same conclusions about evaluability year after year, this looks, in some cases, like a lack of learning. The aid evaluators' own planning and analysis of recurring similar issues and projects ought to have improved more over the years if they had drawn on previous conclusions, circumstances and experiences and tried to learn and develop operations. Many of the findings ought to be inputs to, rather than conclusions of, yet another method chapter.

*If* the conclusion arrived at is that shortcomings regarding knowledge and learning have, over time, affected the reliability of Sida's evaluations, this could, for example, require measures such as:

- External quality assessment of a random sample of evaluations each year carried out by international, independent evaluation expertise.
- Expanded use of research environments, researchers and individuals trained in research in evaluation work.
- Continued joint consultant/framework agreement-wide learning activities focusing on methodology and learning from experience of evaluating aid interventions.
- Tests of knowledge in framework agreement procurements in the area of evaluation produced by international independent evaluation expertise.
- The drafting of a specific quality policy for Sida's evaluations focusing on reliability and methodology. Alternatively, an expansion of the present evaluation handbook.

# References


Alexius, S. (2017), Att styra komplexa uppdrag mot mål och resultat – en forskningsöversikt, Trafikanalys.

Andrews, M., Pritchett, L., Woolcock, M. (2012), *Escaping Capability Traps through Problem-Driven Iterative Adaptation (PDIA*), HKS Faculty Research Working Paper.

Argyris, C. och Schön, D. (1974) Theory in practice: Increasing professional effectiveness, San Francisco: Jossey-Bass.

Bandstein, S. (2006), *Views on Evaluation*, Sida.

Beach, D. and Brun Pedersen, R. (2013), *Process-Tracing Methods. Foundations and Guidelines*. The University of Michigan Press.

Burman, M. and Hårsmar, M. (2015), *Utvärdering av svenskt bistånd – en kartläggning*, Rapport 2015:02 från Expertgruppen för biståndsanalys.

Burman, M. (2017), *Livslängd och livskraft: Vad säger utvärderingar om svenska biståndsinsatsers hållbarhet?,* Expertgruppen för biståndsanalys.

Befani, B. (2021), *Credible explanations of development outcomes: improving quality and rigour of Theory-Based Evaluation with diagnostic approaches and Bayesian Updating*, Forthcoming, Expertgruppen för biståndsanalys.

Befani, B. (2016), *Pathways to Change: Evaluating development interventions with Qualitative Comparative Analysis (QCA),* Expertgruppen för biståndsanalys.

Carlsson, J., Eriksson-Baaz, M., Fallenius, A-M., Lövgren, E. (1999), *Are Evaluations Useful? Cases from Swedish development cooperation*, Sida.

Carlsson, J (1998), *Swedish Aid for Poverty Reduction: A History of Policy and Practice*, ODI.

Carlsson, J., Forss, K., Metell, K., Segnestam, L., Strömberg, T. (1997), *Using the Evaluation Tool. A survey of conventional wisdom and common practice at Sida*. Sida.

Carlsson, J. and Wohlgemuth, L. (1996), *Capacity Building and Networking. A meta-evaluation of African regional research networks*. Sida.

Carlsson, J., Köhlin, G., Ekbom, A. (1994), *The Political Economy of Evaluation. International Aid Agencies and the Effectiveness of Aid*, Palgrave Macmillan UK.

Carlsson, J. and Forss, K. (1997), *The Quest for Quality- Or Can Evaluation Findngs be Trusted?* Evaluation.

Chapman, N., Lloyd, R., Villanger, E., Gleed, G. (2017), *The Quality of Reviews and Decentralised Evaluations in Norwegian Development Cooperation*, Norad.

Chianca, T. (2008)*, The OECD/DAC Criteria for International Development Evaluations: An Assessment and Ideas for Improvement*, Journal of MultiDisciplinary Evaluation.

Christoplos, I., Liljelund-Hedqvist, A. och Rothman, J. (2014), *Lessons and Reflections from 84 Sida Decentralised Evaluations 2013 – a Synthesis Review*, Sida Studies in Evaluation.

Colin, E., Sandström, C., och Wennberg, K. (2021), *Utvärderingar av näringspolitik – en intressekonflikt mellan myndigheter, konsultföretag, politik och skattebetalare?* Ekonomisk debatt, 4/2021.

Copestake, J., Morsink, M., Remnant, F. (2019), *Attributing Development Impact. The qualitative impact protocol case book,* Practical Action Publishing.

Davidson, E.J. (2004), Evaluation Methodology Basics: The Nuts and Bolts of Sound Evaluation. Beverly Hills, CA: Sage Publications.

Davies, R. and Dart, J. (2005) *The 'Most Significant Change' Technique – A Guide to Its Use.*

Dijkstra, G. (2018), *Aid and good governance: Examining aggregate unintended effects of aid*, Evaluation and Program planning.

Doran, G. T. (1981), *There's a SMART way to write management's goals and objectives*. Management Review.

Easterly, W. (2006), *Planners versus Searchers in Foreign Aid.* Asian Development Review.

Easterly, W. (2007), *The White Man's Burden: Why the West's Efforts to Aid the Rest Have Done So Much Ill and So Little Good*, Penguin Group.

Eliasson, K. (2006) *Måluppfyllelseanalys. Hur måluppfyllelse, effekter och effektivitet kan undersökas och rapporteras*, Ekonomistyrningsverket.

Ellström, P-E., (2011), *Informal Learning at Work: Conditions, Processes and Logics*, The SAGE Handbook of Workplace Learning, 2011.

Forss, K. (1994), *Några aspekter på kvaliteten hos utvärderingar av svenskt biståndssamarbete*. Rapport till Sekretariatet för Analys av Utvecklingsamarbete (SASDA).

Forss, K. (2007), Utvärdering som hantverk. Bortom mallar och manualer. Studentlitteratur.

Forss, K., Vedung, E., Kruse, S-E., Mwaiselage, A., Nilsdotter, A. (2008), Are Sida Evaluations good enough? An Assessment of 34 evaluation reports, Sida Studies in Evaluation, 2008:1.

Forss, K., Marra, M. och Schwartz, R. (2011), Evaluating the Complex: Attribution, Contribution, and Beyond, 1st edition, Routledge.

Forss, K. and Pollard, A. (2020), Responses to Decline. The Persistence of Quality Problems, chapter 10 in Perrin, B. and Tyrrell, T. (red.) Changing Bureaucracies, Adapting to Uncertainty, and How Evaluation Can Help, 1st edition, Routledge.

Forss, K. (2021), From Measuring impact to Understanding Change in Long Term Perspectives in Evaluation. Increasing Relevance and Utility, Kim Forss, Ida Lindkvist and Mark McGillivray, Routledge.

Goodhart, C (1975). Problems of Monetary Management: The U.K. Experience. Reserve Bank of Australia.

Holvoet, N., Van Esbroeck, D., Inberg, L., Popelier, L., Peeters, B., Verhofstadt, E. (2018). *To evaluate or not: Evaluability study of 40 interventions of Belgian development cooperation*, Evaluation and Program Planning.

Honig, D. and Gulrajani, N. (2017), Making Good on Donors' Desire to Do Development Differently. Third World Quarterly.

Hood, C. (1991), A public management for all seasons? Public Policy and Administration.

Hydén, G. (2019), Democracy in African Governance: Seeing and Doing It Differently, Expertgruppen för biståndsanalys, 2019.

Isaksson, A-S. (2017), Geospatial analysis of aid: A new approach to aid evaluation, Expertgruppen för biståndsanalys.

Jacobsson, B. and Sundström, G. (2001), Resultat utan lärande: erfarenheter från tre decennier av resultatstyrning, Stockholms centrum för forskning om offentlig sektor (SCORE).

Johansson, L. och Lindgren, M. (2007) Sida Evaluations and Audits 2006, A synthesis of findings and conclusions, Sida.

John W. Evers (1980), A Field Study of Goal-Based and Goal-Free Evaluation Techniques. Western Michigan University.

Karlsson Vestman, O. (2011), Utvärderandets konst., Studentlitteratur.

Karlström, U., Eriksson, C., Wockelberg-Hedlund, J. (2018), Analyser och utvärderingar för effektiv styrning. Report of the Inquiry on the Government's analysis and evaluation resources, SOU 2018:79.

King, G., Keohane, R. and Verba, S. (1994), Designing Social Inquiry. Scientific Inference in Qualitative Research, Princeton Univrsity Press.

Krogstrup, H-K. (2017), Utvärderingsmodeller, Studentlitteratur.

LaGuardia, D., Becker-Thierry, S., Bonoll, F. (2014), Mid Term Review of the Framework Agreement for Sida Reviews, Evaluations and Advisory Services on Results Frameworks Final Report, Sida Decentralised Evaluation, 2014:13.

Leeuw, F. (2003), Reconstructing Program Theories: Methods Available and Problems to be Solved, American Journal of Evaluation.

Lindblom, C. (1959), The Science of 'muddling through.' Public Administration Review.

Lindgren, L. (2012), Terminologihandbok för utvärdering, Studentlitteratur.

Mayne, J. (2012), Contribution analysis: Coming of age?, Evaluation.

Molander, P. (2017), Dags för omprövning – en ESO-rapport om styrning av offentlig verksamhet, ESO-rapport 2017:1.

Mosley, P. (1986). Aid-effectiveness: The micro-macro paradox. Institute of Development Studies Bulletin.

Nielsen, S-E., Lemire, S., Christie, C-A. (2019) The Commercial Side of Evaluation. Evaluation as an industry and as a Professional Service. I. Furubo, J-E. and Stame, N., The Evaluation Enterprise. A Critical View. Routledge.

Nilén, H., Nilsson, P-U., Nyberg, J. (2002), Syntes av svenska enskilda organisationers utvärderingar, Sida.

Nilsson, K. och Widell, L. (2006), Effektutvärdering. Att välja upplägg, Ekonomistyrningsverket.

OECD/DAC (2019), Better Criteria for Better Evaluation. Revised Evaluation Criteria Definitions and Principles for Use, OECD/DAC Network on Development Evaluation, 10 December, 2019.

OECD/DAC (2016), Evaluation Systems in Development Co-operation. 2016 Review, OECD/DAC.

Ostrom, E., Gibson, C., Shivakumar, S., Andersson, K. (2001), Aid, Incentives, and Sustainability. An Institutional Analysis of Development Cooperation, Sida Studies in Evaluation, 2002:01.

Pain, A. (2015), Review of Sida's Support to Afghanistan – Lessons and Conclusions from 7 Evaluations, Final Report, Sida.

Patton, M. Q. (2008), Utilization-Focused Evaluation, Fourth Edition, Sage Publication.

Patton, M. Q. (2018). Principles-Focused Evaluation – The GUIDE. New York: Guilford Press.

Pawson, R. och Tilley, N. (1997), Realistic Evaluation. London: Sage.

Pawson, R. (2013), The Science of Evaluation. A Realist Manifesto. Sage Publication Ltd.

Peck, L., Jassey, K., Johansson, J. (2006), Sida Evaluations and Audits 2005, Sida.

Peck, L. and Knippel, V. (2008), Sida Evaluations and Audits, 2007. A Synthesis of Findings and Conclusions. UTV Working Paper, Sida.

Peck, L. (2007), Erfarenheter av resultatstyrning. En genomgång av utvärderingar och studier, Sida.

Petersson and Sandahl (2016), Kausalitet i filosofi, politik och utvärdering, Studentlitteratur.

Pleger, L., Sager, F., Morris, M., Meyer, W., Stockmann, R. (2016), *Are Some Countries More Prone to Pressure Evaluators Than Others? Comparing Findings from the United States, United Kingdom, Germany, and Switzerland.* American Journal of Evaluation.

Pritchett, L. (2002), It pays to be ignorant: A simple political economy of rigorous program evaluation, The Journal of Policy Reform.

Riddell, R., Bebbington, A., Peck, L. (1995), Promoting Development by Proxy. An Evaluation of the Development Impact of Government Support to Swedish NGOs, Sida Evaluation Report, 1995:02.

Riddell, R. (2007), Does Foreign Aid Really Work?, Oxford University Press.

Riksrevisionsverket (1998), Lär sig SIDA? En granskning av SIDAs förmåga att lära sig av erfarenheterna: revisionsrapport, Riksrevisionsverket.

Rogers, P. och Weiss, C. (2007), Theory-Based Evaluation: Reflections Ten Years On. Theory-Based Evaluation: Past, Present, and Future. New Directions for Evaluation.

Samoff, J. (2021), Relevant? Almost always. Relevance i Development Cooperation. Working Paper. Expertgruppen för biståndsanalys.

Samoff, J., Leer, J., Reddy, M. (2016), Capturing Complexity and Context: Evaluating Aid to Education, Expertgruppen för biståndsanalys.

Sandahl, R. (2019), Utvärdering – en balansakt, PM publicerat av Tillväxtverket.

Sandberg, B. och Faugert, S. (2012), Perspektiv på utvärdering, Upplaga 2, Studentlitteratur.

Scriven, M. (1991), Pros and Cons about Goal-Free Evaluation, Evaluation Practice, 1991.

Schwadt, T.A. (2019) The concerted effort to Professionalize Evaluation Practise. Whiter Are We Bound? The Evaluation Enterprise. A Critical View. Edited by Jan-Eric Furubo and Nicoletta Stame. Routledge.

Shutt, C. (2016), Towards an Alternative Development Management Paradigm?, Expertgruppen För biståndsanalys.

Stame, N. (2004), Theory based Evaluation and Types of Complexity, Evaluation.

Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., Befani, B. (2012) Broadening the Range of Designs and Methods for Impact Evaluations, Working paper 38, Department for International Development UK.

Sundström, G. (2001) ”Ett relativt blygsamt förslag” – Resultatstyrningens framväxt ur tre perspektiv. Score Rapportserie 2001:2.

Tarschys, D. (2006), Mål utan mening? Om ordstyrning i landet där allt är prioriterat, i Förvaltningens byggstenar – rapport från ett seminarium, Statskontorets småskrifter, Statskontoret, 2006.

Tobisson, E. och De Vylder, S. (1997), Poverty Reduction and Gender Equality. An Assessment of Sida´s Contry Reports and Evaluations in 1995–96. Sida.

Uggla, F. och Wulfing, M. (2007), Assessment of Sida's Support for Human Rights and Democracy. Sida.

Vaessen, J., Lemire, S. och Befani, B. (2020), Evaluation of International Development Interventions. An Overview of Approches and Methods. Independent Evaluation Group.

Vedung, E. (2004), Utvärderingsböljans former och drivkrafter, FinSoc Working Paper.

Vedung, E. (2012), Utvärdering i politik och förvaltning, tredje upplagan, Studentlitteratur.

Vedung, E. (2016), Implementering i politik och förvaltning, Studentlitteratur.

Vähämäki, J. (2017), Matrixing Aid: The Rise and Fall of 'Results Initiatives' in Swedish Development Aid, Stockholms universitet.

Wilson-Grau, R. (2018), Outcome Harvesting: Principles, Steps, and Evaluation Applications., Information Age Publishing.

## Official publications and websites


Blog about the terms contribution and attribution in aid evaluation: https://aidleap.org/2015/11/23/contribution-vs-attribution-a-pointless-debate/ (read 18 December 2020).

*Cochrane Handbook for Systematic Reviews of Interventions*, Version 6.2, 2021 Cochrane Training, https://training.cochrane.org/handbook/current

Om Contribution Analysis: https://www.betterevaluation.org/en/plan/approach/contribution_analysis

Ray Pawson on complexity and time in evaluation: https://www.youtube.com/watch?v=c8TReTAbEQw&feature=youtu.be

Regeringens Policyramverk för svenskt utvecklingssamarbete och humanitärt bistånd, Regeringens skrivelse 2016/17:60.

Resultatutvärdering- ett programförslag, SIDA 1971.

Sida's Evaluation Handbook. Guidelines and Manual for Conducting Evaluations at Sida. 2018, Sida.

Strategi för stöd genom svenska organisationer i det civila samhället, 2016–2022, Utrikesdepartementet, 2016.

# Annex 1. Implementation and method

The study has been implemented in five stages:

1. Review of previous studies focusing on effectiveness and the analysis of effectiveness in and outside Swedish development assistance.
2. Download of a sample of evaluations (decentralised and strategic) from Sida's publication database.
3. Drafting and testing of the analytical framework.
4. Review of the evaluations based on the framework.
5. Compilation and analysis of data and writing of the report.

The implementation process is described and discussed in detail below (stage by stage as above).

The method used in the report is a development of the method used in *Livslängd och livskraft: Vad säger utvärderingar om svenska biståndsinsatsers hållbarhet?* (EBA 2017:12), focusing on the OECD/DAC evaluation criterion of Sustainability.

The whole of the work was implemented in the period May 2019 to April 2021.

## 1. Review of previous studies

The work began with a review of:

1. Literature on objective-based evaluation.
2. Studies of effectiveness in Swedish aid (primarily those focused on several interventions at the same time or that synthesised evaluations) and
3. Previous studies of Sida's evaluation work and the quality of Sida's evaluations.

Points 2 and 3 above refer almost solely to 'grey literature' and literature not published in scientific journals. Many of these studies were commissioned by Sida itself. The review is reported in chapters 2 and 3.

The review of the literature (point 1 above) does not claim to capture all of the literature about evaluation and effectiveness, even though the bulk of what is available in book form in Swedish has been reviewed (see references).

The purpose of this review of previous studies has been to establish a simple theoretical foundation for the analytical framework and to place the conclusions of this study in a longer and broader perspective. The latter purpose means that we do not intend to strictly apply all the knowledge from the theoretical part to the empirical material.

The review of the literature has been important since this study should be viewed in a broader context in which donors have, over time, repeatedly commissioned studies that have summed up conclusions from evaluations, analysed the quality of the evaluations and then (as it turns out) drawn similar conclusions.

Literature searches have been conducted using google and google scholar, in Sida's publication database, on Openaid.se, on biståndsdebatten.se/fuf.se and via lists of references in literature already known to the author.

## 2. Sample and downloading of evaluations

The decentralised evaluations in the sample were downloaded on 24 January 2020 from Sida's publication database and the strategic evaluations were downloaded on 7 February 2020. All decentralised and strategic evaluations that were available at that time in Sida's publication database for the years concerned were downloaded. A total of 98 decentralised evaluations for 2017 (35), 2018 (30) and

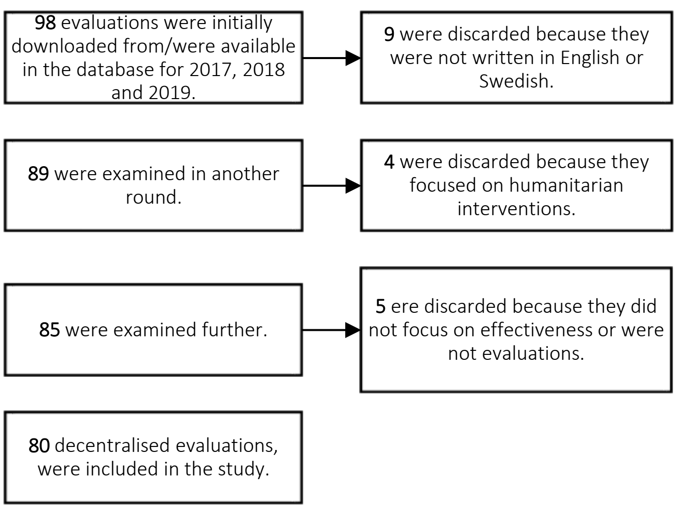
2019 (33) and 8 strategic evaluations for 2015 (2), 2016 (2), 2017 (1), 2018 (2) and 2019 (1).

Then reports were discarded that 1) were not written in Swedish or English, 2) did not focus on effectiveness, 3) focused on humanitarian aid and 4) turned out not to be evaluations.

There then remained a total of 80 decentralised evaluations and 6 strategic evaluations as the basis for the study.

Figure B1:1 describes how the sample of 80 studies was obtained.

**Figure B1:1: Sample of decentralised evaluations, from 98 to 80**




Source: The author.

Out of the decentralised evaluations, 35 reports had been published in 2017, 29 in 2018 and 26 in 2019.

Of the initial 9 *strategic evaluations*, three reports were removed. Sida 2016:2 was removed because it focused on Sida's intervention management rather than the intervention as such. The evaluation of the Market Systems Development approach (MSD) did not analyse effectiveness and was removed. The Annual Report for 2017 mentions an evaluation of the interaction between humanitarian and long-term development cooperation, but it does not seem to have been published and is not included either.[75]

## Sample and representativeness

Slightly simplified, Swedish aid consists of four large areas; 1) bilateral long-term development cooperation (including 'Mult-bi'), 2) humanitarian aid, 3) multilateral aid (including e.g. EU aid) and 4) various costs in other expenditure areas that are included in the aid 'framework'. The studies in our sampler focus solely on area 1 above and the long-term development cooperation handled by Sida. The other parts (2–4) are not captured by the evaluations we have reviewed.

Nor can the sample be viewed as representative of Sida's long-term development cooperation as a whole in the years studied in the report. 2015–2019 for strategic evaluations and 2017–2019 for decentralised evaluations. Sida's Evaluation Handbook states that when there is a clear use and it is appropriate to do evaluations, for example:

- "For a project or programme with unknown or disputed outcomes.
- For large and expensive interventions.
- For pilot initiatives or innovative projects or programmes.
- When Sida has a strategic interest.
- When stakeholders are keen on an evaluation."

Previous studies (EBA 2015:02 and EBA 2017:12) have also shown that health interventions are seldom evaluated, while research interventions are evaluated fairly often, as are interventions in

---

[75] The report is viewed according to Sida as an internal product, available but not published.

East and Central Europe. Three large areas in our sample are National, regional or local administration (14/80 evaluations), Market, business, commerce, innovation (12/80 evaluations), Research and higher education (10/80 evaluations) and Human rights (8/80 evaluations). Two evaluations focus on gender equality and two evaluations on SRHR.

Geographically, Sub- Saharan Africa is the largest region with 37/80 evaluations, and the second largest region is East and Central Europe with 15/80 evaluations (19 %).[76]

Sida's Evaluation Unit says that all reports are posted on the website. However, it is not possible to rule out there being individual evaluations that were not published for some reason and were therefore missed in the study. Sida in Stockholm is not always aware of embassy procurements and studies may be missing on grounds of secrecy.

In the sample of decentralised evaluations, nine evaluations have been removed because the were not written in English or Swedish. There is no real reason to believe that these reports would have differed greatly in terms of reliability or their assessment of effectiveness. The evaluators are taken from the same consultancies and the same framework agreement procurement, and several of the authors are also found in the English-language reports in the sample. The review shows small differences in conclusions about effectiveness between regions.

The overall conclusion about representativeness is that even though the evaluation sample should not be seen as representative of Sida's intervention portfolio as a whole the sample does reflect Sida's evaluation work (but not humanitarian interventions) in a comprehensive way for the period 2017 to 2019.

[76] Part of the reason why Latin America only has two evaluations is that studies in Spanish and Portuguese have been removed as only three such studies were published in the period included here.

## 3. Development and testing of the analysis framework

An analysis framework was developed on the basis of an initial reading of the literature (reported in chapters 2 and 6 of the main report), and it has steered and given a systematic structure to the review (annex 3). Parts of the framework (the background variables especially) are a developed and improved version of the framework used in EBA 2017:12.

The starting point for the report has been OECD DAC's previous definition of effectiveness for aid evaluations (see chapter 1 and annex 4). OECD DAC's new definition from 2019 has not been used since the evaluations in the sample were ordered and carried out before the change. There is some discussion of the new definition of the effectiveness criterion in annex 4.

The framework was first tested and developed using a random sample of eight evaluations from the gross sample and has since been developed gradually during the course of the work.[77] The starting point was to arrive at clear and simple questions, where it was judged to be relevant and possible to look for answers in the report. Each scale stage has been specified for the question of effectiveness.

## 4. Review of the evaluations based on the framework

After that the evaluations were reviewed systematically on the basis of the analysis framework. Basic data from the review have been compiled and documented in an Excel database that is available from the EBA. The assessment was replicated for 80 randomly selected observations spread across the data set to check reproducibility. Chapter 4 discusses a couple of factors that may have partly reduced

[77] All randomisation is based on figures from www.Random.org

reproducibility in the synthesis of the effectiveness question: the fact that the consultants' assessments are often unclear, presented in indirect terns and complex and that overall assessments are not always made Around 10 % of the evaluations (nine decentralised and one strategic) had such unclear conclusions that the scale stage of the assessments could not be determined.

One conclusion after EBA 2017:12 (see EBA 2021:01) was that a synthesis is not fully reproducible unless specifications are given of how the review was conducted for every single question and of how the question has been interpreted specifically by the person doing the review. It has been necessary to further specify numerous questions when different interpretations of the textt turned out to be possible. The degree of reproducibility therefore differs between questions in the framework. The evaluations are quality-oriented and, not seldom, fairly unclear texts. Therefore the study cannot be compared with a systematic review[78] in terms of the systematic approach and process to ensure reproducibility and validity. The most important weakness is that the review has been carried out by one person and not by several persons, as is the rule in a systematic review, for example. It is also of central importance that the answer to question 1 in the review (effectiveness) is viewed in the light of question 2 (reliability).

The study has only analysed published evaluation reports. If the evaluators have described their work in an inception report not attached, this is not picked up by the analysis. Several reports did not have an inception report. This is judged to play a minor role for the conclusions of the analysis about reliability since it is fundamental quality practice that the description of the method is set out in the report or an annex.

[78] Cochrane Handbook for Systematic Reviews of Interventions | Cochrane Training.

# 5. Compilation and analysis of the data and writing of the report

The final stage of the work was to analyse and compile the data with the aid of Excel and to write the report with the review of previous studies and evaluations as background.

The work has been followed by a reference group that commented the initial project proposal the analysis framework and the drafts of the report. The fact that one member of the reference group is employed at Sida's Evaluation Unit and that one reference group member works as an evaluation consultant for Sida may have influenced the discussion in the reference group.

It should be pointed out that, in practice, the five stages (above) have overlapped. For example, chapters 1, 2 and 6 were being worked on as drafts while the review of the evaluations was in progress, In this way,the study has evolved in an interplay between reading previous studies in the area and the review of the evaluations.

# Annex 2. Analysis framework

The analysis framework's questions are listed below with response alternatives, specifications, limitations and a rough estimate of confidence per question (the confidence is: very good, good, moderate or low).[79]

## 1. Title of the evaluation

Specification and limitations: open-ended responses, confidence: very good

## 2. Number of the evaluation

Specification and limitations: 1–80, specification: random assignment based on random number generator, confidence: very good

## 3. Country

Specification and limitations: open-ended responses, all countries studied in the evaluation stated, confidence: very good

## 4. Region

Sub-Saharan Africa
Asia, incl. Central Asia
Central and Eastern Europe
Latin America
MENA
Global

Specification and limitations: based on question 3 above, global corresponds to more than one region, confidence: very good

[79] The degree of reproducibility probably decreases with the level of confidence.

## 5. Geographical focus of evaluation object

Country or sub-national level

Region (more than one country in a region)
Global (several regions)

| |
|---|
| Specification and limitations: based on question 3 above, global corresponds to more than one region, region corresponds to more than one country, confidence: very good |

## 6. Period of time evaluated

| |
|---|
| Specification and limitations: open-ended responses, number of whole years, sometimes difficult to see in the event of several interventions and borderline cases, rounded to whole years, confidence: good |

## 7. Are several interventions analysed at the same time?

One intervention
Several at the same time

| |
|---|
| Specification and limitations: number of evaluation objects in the evaluation (projects, programmes, etc.), no problem to see, confidence: very good |

## 8. The evaluation object's total number of years with Sida funding at the time of the evaluation

1–4 years
5 years or more
Not stated in the report/cannot be seen/not relevant

| |
|---|
| Specification and limitations: refers to number of years of funding of the activity at the time of the evaluation irrespective of whether project or programme, not, for example, specific objectives that may have been changed over time, see also question 6. Confidence: good |

## 9. Object of the evaluation

Programme
Project
Organisation
Strategy or plan
Guarantee
Fund
Unclear or other

| |
|---|
| Specification and limitations: Completely follows the categorisation/designation of the evaluation object used by the evaluator in the report, confidence: very good |

## 10. Implementing partner

Public (excl. multi)
Civil society
Private sector
Multilateral organisation, development bank, intergovernmental organisation
Cross-border cooperation (several of the above)

| |
|---|
| Specification and limitations: Form/type of organisation running or implementing the intervention, organisations that are not public entities, private companies, development banks or multilateral organisations have been classed as civil society, confidence: very good |

## 11. What questions are to be answered under the effectiveness criterion according to the evaluation's terms of reference?

| |
|---|
| Specification and limitations: open-ended responses taken from reports, all evaluation questions in the terms of reference that affect the categories in question 13, plus results, effectiveness and impact, component of personal assessment regarding what is associated with the efficiency criterion, confidence: good |

**12. Do the terms of reference ask explicitly for an assessment of effectiveness that does not focus on overall or long-term objectives?**

Yes

No

Unclear/text contradictory in this respect

Specification and limitations: the question only applies to evaluations that have studied a single intervention. Refers only to what is said explicitly/expressly in the terms of reference. There may be a limitation in the ambiguity entailed by the terns outcome, impact and effectiveness. More specifically formulated terms of reference may also describe exactly what results are to be followed up and have no reason to specify the level of objectives explicitly. The use of language has been unclear. Confidence: good

**13. Do the terms of reference ask explicitly for an assessment of the intervention's effectiveness or are other terms used in formulating the evaluation questions?**

Yes (an assessment of effectiveness is asked for)

No (an assessment of effectiveness is not asked for explicitly; instead they stress expected results or specify particular results that are to be the focus of the assessment)

Specification and limitations: refers to what is said explicitly/expressly in the terms of reference, however there may be borderline cases, confidence: good

**14. What do the terms of reference ask for in addition to the reply their fundamental question (effectiveness/expected results)? (Several responses possible)**

Analysis regarding:

Specific types or break-downs of results/effectiveness
Mechanisms/explanations (of effectiveness/results)

Programme theory/theory of change/LFA
The intervention's result and monitoring work or indicators
Side effects, negative or unforeseen effects.
Can the objectives be reached in the next phase, going forward/in the future?
What can be done to reach the objectives?
Whether the objectives or expected results are clearly stated
Have various outputs or results been delivered in time?

Nothing else
Something else, state what:

Specification and limitations: the focus is wholly on the evaluation questions enumerated in the terms of reference. Component of personal assessment regarding what response alternatives to use. The categorisation should, however, have contained more response alternatives, and there is a small risk that issues concerning effectiveness have been missed because they have been categorised under other evaluation criteria, confidence: good

## 15. Have the interventions objectives/expected results have been changed or adjusted during the period in focus according to the evaluation?

Yes, they have been reconstructed, clarified or restructured by the evaluators.

Yes, but primarily for other reasons (for example at transitions between project periods)

No, probably not

The evaluation focuses on more than one intervention (is therefore not analysed)

Not possible to see from the report/not stated

Specification and limitations: partly difficult to see from the reports, partly vague language use in the reports. It is possible that the objectives have been stated in more detail or changed without this being mentioned by the authors. Confidence: moderate

## 16. What development stage in terms of time is highest in the focus on effectiveness/the results analysis*?*

Output
Outcome
Impact
Not possible to see

| |
|---|
| Specification and limitations: refers to what the evaluators themselves consider/state is in the focus of the analysis in the evaluation text. Not necessarily what was actually analysed. There is a limitation here in the confusion about terms in the area. Confidence: good |

## 17. Are the expected long-term results of the intervention to be achieved directly in society or nature or in processes or activities?

In processes, organisations, activities that are to be strengthened, built or changed (that are then able to directly/indirectly influence people/society/nature)

Externally in the communities or societies affected (directly in relation to citizens, rights holders, in nature or ecosystems)

I both and

Not relevant/cannot be seen

| |
|---|
| Specification and limitations: the conclusion here us based on the stated objectives/expected results and not on what was evaluated in practice. There has, however, been some interpretation and borderline cases. Confidence: good |

**18. Which of the following categories correspond to the overall long-term objectives of the intervention? (Several alternatives possible)**

Change or improve the functioning or ways of working of public institutions or capacity in a sector

Strengthen implementation of or compliance with human rights in one or more areas (thematically or geographically)

Strengthen capacity and conduct research

Strengthen the use of research results

Create mobilisation or participation or generate demand for democracy

Reduced poverty in a region, country or part of a country

Economic empowerment for a particular group

Conflict resolution, peace, peaceful societies and security

Spread knowledge, train, develop skills, inform about an area/issue

New and better jobs/job opportunities

Better living conditions or higher incomes

Strengthen commerce

Reduce violence in various forms or contexts or in relation to different groups

| |
|---|
| Specification and limitations: the purpose is to use a generic approach to illustrate over levels of objectives. Another author might have formulated the categories differently. Only categories with three or more occurrences are reported. Confidence: moderate |

## 19. Do the evaluators discuss, appraise or analyse how the intervention's objectives or expected results have been formulated (in terms of, for example ambition, degree of specification, measurability, realism, acceptance or time perspectives)?

Yes

No

If yes, give examples of what is said: ... (Open-ended response, enter the most important conclusions)

Specification and limitations: The question concerns the wording and the objectives' form and not substance. The question here only picks up explicit discussions of objectives/expected results and not the logical framework or the like. In some cases the evaluators may therefore have wanted to express a tacit comment on the wording of objectives, but this was not picked up Some evaluations also focus on more than one intervention. Confidence: moderate

## 20. To what extent have objectives or expected results been achieved so far in the intervention?

| Scale stages | Criteria |
|---|---|
| Wholly fulfilled (objectives/results) | • With the exception of one or a few single, or marginal, objectives/results, the assessment is that all objectives/results are fulfilled or are on the way to being reached/will be reached, and/or:<br>• The intervention is described as being effective almost without any exceptions. |
| Mainly fulfilled (objectives/results) | • The evaluator makes the assessment that the majority of objectives/expected results (or the objectives that the evaluator has stated as most important) are fulfilled or on the way to being reached/will be reached, and/or: |

| Scale stages | Criteria |
|---|---|
| | • The intervention is described as mainly effective |
| Mainly not fulfilled (objectives/results) | • The evaluator makes the assessment that the majority of objectives/expected results (or the objectives that the evaluator has stated as most important) are not fulfilled or on the way to being reached/will be reached, and/or<br>• The intervention is described as being mainly ineffective |
| Not fulfilled (objectives/results) | • No or only a few objectives/results, or objectives/results of marginal importance, are judged by the evaluator to have been achieved or to be on the way to being reached/will be reached, and/or:<br>• The intervention is described as ineffective almost without any exception. |
| Not possible to see in the report/conclusions to unclear for determination of scale stage. | • Information wanted cannot be seen in the report.<br>• The conclusions are very unclear.<br>• Or the evaluator has not made any assessment (despite being commissioned to do so) |

Specification and limitations: The conclusions are often unclear or complex and the studies seldom conclude clearly. Confidence has been partly strengthened by the fact that the criteria for the individual scale stages that make up a four-stage scale corresponding to 25 % each in the distribution. The criteria have steered the review, not the names of the scale stages. The reproducibility is probably lower in borderline cases between scale stages. Confidence: moderate

## 21. Do the evaluators describe specifically and fully how the effectiveness question has been answered. (%)

Yes

No

Specification and limitations: What is specific and clear is partly a matter of judgment. The focus has been on the inception report and the method chapter with a special focus on any evaluation matrix. The starting points have been to find yes rather than no in assessment of borderline cases. Confidence: moderate

## 22. What evaluation design does the evaluator refer to in their effectiveness analysis? (%)

Contribution analysis
Theory based evaluation (general)
Outcome harvesting
Outcome mapping
Most significant change
Process tracing
Realist evaluation
Systems based Evaluation
QCA (Qualitative Comparative Analysis)
QUIP (The Qualitative Impact Assessment Protocol)
Regression-based analysis (some form)
Quasi-experiment (some form)
Randomised control study
Combination of two alternatives
Other/Unclear/None

Specification and limitations: The analysis was made deductively using a search function. Only design alternatives mentioned explicitly and clearly in the text are picked up. One weakness is the ability of the analysis to pick up other design alternatives not listed. Confidence: very good

## 23. Does the evaluator test a specified theory of change against the empirical data in the analysis of effectiveness?

Yes

No

Unclear/cannot be seen

| |
|---|
| Specification and limitations: The assessment only picks up what is said explicitly in the report's method chapter and inception report. Some reports did not have an inception report. In borderline cases we have chosen to conclude yes rather than no. Confidence: good |

## 24. Do the evaluators describe how they assess and have tried to deal with causality and potential external influencing factors in their analysis of effectiveness? (%)

Yes

No

Unclear/cannot be seen

| |
|---|
| Specification and limitations: The assessment only picked up what has been said explicitly in the report's method chapter and inception report. Some reports did not have an inception report In borderline cases we have chosen to conclude yes rather than no. Confidence: good |

## 25. Are the terms counterfactuality, additionality or causality found in the evaluation?

Yes + which

No

| |
|---|
| Specification and limitations: The analysis was carried out using the search function in Adobe based on the terms' root and different endings. Confidence: very good |

## 26. Does the evaluation's method chapter take up any methodological limitations specifically concerning the assessment of the effectiveness criterion and/or causality?

Yes for effectiveness and for causality
Yes for effectiveness but not for causality
Yes for causality but not for effectiveness
No
The report's method chapter does not have section on limitations.
If yes, what is said:

| |
|---|
| Specification and limitations: The question focuses on what is said explicitly under the heading Limitations in the method chapter. Virtually all evaluations did have such a heading. At the same time, assessing which limitations touch on effectiveness has involved interpretation. Reproducibility in relation to the question of causality is judged to be slightly better than reproducibility in relation to effectiveness in general. Confidence: moderate |

## 27. According to its terms of reference, is the evaluation to analyse or discuss (negative or positive) unforeseen results or effects?

Yes

No

| |
|---|
| Specification and limitations: The question focuses on what is asked for explicitly in the report's terms of reference. The evaluation may very well have focused on unforeseen outcomes or impacts without them being asked for there. Confidence: very good |

# Annex 3. Evaluations in the sample

## Decentralised evaluations (80)

### 2017

1. *Midterm Review of GESTERRA Capacity Building Programme on Land Management and Administration within DINAT*. Final Report. Ian Christoplos, Dale Doré, André Calengo, Svend Erik Sørensen. Sida Decentralised Evaluation 2016:15.
2. *Evaluation of Sida's Support to the Gorée Institute (2012–2015). Final Report*. Jérôme Gouzou, Olivia Lazard. Sida Decentralised Evaluation 2017:1.
3. *Evaluation of Sida's Support to Tostan (2010–2016). Final Report*. Kimiko Hibri Pedersen, Alassane Binta Diop, Helena Neves, Olivia Lazard, Jérôme Gouzou. Sida Decentralised Evaluation 2017:2.
4. *Evaluation of the Sida supported RFSL projects "LGBT Voices" and "Rainbow Leaders"*. Final Report. Annika Nilsson, Jessica Rothman. Sida Decentralised Evaluation 2017:4.
5. *Final Evaluation of the Natural Resource Management Facility At Act!* Final Report. Eric Buhl-Nielsen, Winnie Wangari Wairimu. Sida Decentralised Evaluation 2017:6.
6. *Evaluation of Sida Support to ECPAT International*. Final Report. Louise Mailloux, Agneta Gunnarsson. Sida Decentralised Evaluation 2017:5.

7. *Evaluation of Swedish government research cooperation with Eduardo Mondlane University, Mozambique 2011–2016*. Synthesis Report. Stein-Erik Kruse, Inge Tvedten, Matti Tedre, Carmeliza Soares da Costa Rosário. Sida Decentralised Evaluation 2017:9.
8. *Evaluation of the OECD's 'Sector Competitiveness Strategy' Project in Ukraine. Final Report.* Andrea Spear, Vera Devine, Oleh Myroshnichenko. Sida Decentralised Evaluation 2017:10.
9. *Mid-Term Evaluation of the National Employment Programme, Rwanda.* Final Report. Sarah Gray, Charles Twesigye-Bakwatsa, David Muganwa, Marcienne Umubyeyi. Sida Decentralised Evaluation 2017:11.
10. *Evaluation of the Regional Statistics Cooperation on the Western Balkans 2013–2016*. Final Report. Leif Danielsson, Sabina Ymeri, Pier Giorgio Ardeni. Sida Decentralised Evaluation 2017:12.
11. *Evaluation of the Sida-funded Partnership in Statistics: A cooperation project between Statistical Office of the Republic of Serbia (SORS) and Statistics Sweden (SCB)*. Final Report. Pier Giorgio Ardeni, Leif Danielsson, Sabina Ymeri. Sida Decentralised Evaluation 2017:13.
12. *Evaluation of Core Support to Georgian Foundation for Strategic and International Studies (GFSIS)*. Final Report. Pierre Walther, Nelly Dolidze. Sida Decentralised Evaluation 2017:15.
13. *Evaluation of the Health Economics and HIV and AIDS Research Division (HEARD)*. Final Report. Silke Graeser, Anne L. Buffardi, Lennie Bazira Kyomuhangi. Sida Decentralised Evaluation 2017:17.
14. *Completion Evaluation of Sida Support to Environment and Climate Change Component of NREP*. Final Report. Leif Danielsson, Denis Rugege, Richard Ngendahayo. Sida Decentralised Evaluation 2017:18.

15. *Review of the Statistical Cooperation project between the National Institute of Statistics of Albania and Statistics Sweden, Phase IV, October 2014 – December 2017*. Final Report. Leif Danielsson, Sabina Ymeri. Sida Decentralised Evaluation 2017:20.

16. *Evaluation of Swedish Support to the One UN in Albania for gender equality work 2012–2017*. Final Report. Jonas Lövkrona, Sabina Ymeri. Sida Decentralised Evaluation 2017:21.

17. *Mid Term Review of the Agadir Technical Unit and the Swedish International Development Agency, Sida funded project "Support Quality Infrastructure in Agadir Countries"*. Final Report. Ali Dastgeer, Bourhan Kreitem, Jens Andersson, Stefano Ciarli. Sida Decentralised Evaluation 2017:22.

18. *Evaluation of ITP 299 Strategies for Chemicals Management*. Final Report. Ali Dastgeer, Abigail Hansen, Adriana Jalba, Alicia Borges Månsson. Sida Decentralised Evaluation 2017:23.

19. *Evaluation of Sida's research cooperation with Bolivia, 2007–2016*. Synthesis Report. Ananda S. Millard, Matti Tedre, Erik W. Thulstrup, Manuel Ramiro Munoz, Pamela Velasco. Sida Decentralised Evaluation 2017:24.

20. *End of Strategy Evaluation of the Zanzibar Legal Services Centre. Final Report*. Greg Moran, Clarence Kipobota, Johanna Lindgren Garcia. Sida Decentralised Evaluation 2017:26.

21. *Evaluation of EuroMed Rights Promotion of Human Rights in the Middle East and North Africa*. Final Report. Jérôme Gouzou, Kimiko Hibri Pedersen, Kristoffer Engstrand. Sida Decentralised Evaluation 2017:27.

22. *Evaluation of three projects on transboundary water management in the Middle East and North Africa region. Final Report*. Nicklas Svensson, Ludo Prins, Aly Kerdany. Sida Decentralised Evaluation 2017:28.

23. *End-Term Evaluation of the Swedish Education Support to Zanzibar 2010–2017*. Final Report. Sheila Reed, Angela Arnott, Idrissa Yussuf Hamad, Kristeen Oberlander Chachage. Sida Decentralised Evaluation 2017:32.
24. *End-Term Evaluation of the Global Partnership for Education (2014–2016) in Zanzibar*. Final Report. Sheila Reed, Angela Arnott, Idrissa Yussuf Hamad, Kristeen Oberlander Chachage.
25. *Evaluation of 3rd Call off of civil society support through umbrella organisations 2013–2017*. Final Report. Kathleen Webb, Agneta Gunnarsson, Camille Pellerin, Abdulatif Idris, Bezawit Bekele. Sida Decentralised Evaluation 2017:34.

## 2018

1. *Evaluation of the Sida-USAID/DCA Guarantee to Zanaco*. Final Report. Sarah Gray Emma Sitambuli, Jens Albråten. Sida Decentralised Evaluation 2018:29.
2. *Review of Ratmalana/Moratuwa and Ja-Ela/Ekala Wastewater Disposal Project in Sri Lanka*. Final Report. Åke Nilsson, Arumugham P.K., Nalin Wikramanayake, Kusum Athukorala. Sida Decentralised Evaluation 2018:30.
3. *Evaluation of the National Legal Aid Clinic for Women's Access to Justice Programme in Zambia*. Final Report. Greg Moran, Dr. Lungowe Matakala, Lovisa Arlid, Johanna Lindgren Garcia. Sida Decentralised Evaluation 2018:1.
4. *Evaluation of the Sida supported research capacity and higher education development program in Rwanda, 2013–2017*. Final Report. Inge Tvedten, Arthur Byabagambi, Johanna Lindström, Matti Tedre. Sida Decentralised Evaluation 2018:3.

5. *Evaluation of CIVICUS – World Alliance for Citizen Participation.* Final Report. Bente Topsøe-Jensen, Ingrid Obery, Khilesh Chaturvedi, Raphaëlle Bisiaux. Sida Decentralised Evaluation 2018:4.
6. *Evaluation of Programme Work Methods of The Raoul Wallenberg Institute of Human Rights and Humanitarian Law.* Final Report. Nicklas Svensson, Lone Lindholt, Henrik Alffram. Sida Decentralised Evaluation 2018:5.
7. *Evaluation of Afrobarometer's Regional Programme 2011–2017.* Final Report. Ian Christoplos, Peter da Costa, Dena Lomofsky, Faten Aggad, Linnea Hincks, Raphaëlle Bisiaux. Sida Decentralised Evaluation 2018:6.
8. *Evaluation of the Sida supported programmes "CLACSO Knowledge for a Sustainable World 2013–2016" and "South-South Tricontinental 2013–2015".* Final Report. Ian Christoplos, Francisco Sagasti, Mario Bazan. Sida Decentralised Evaluation 2018:7.
9. *Mid-term Evaluation of Swedish government funded Civil Society Support through the AGIR II Programme in Mozambique 2014–2020.* Final Report. Stein-Erik Kruse, Inge Tvedten, José Jamie Macuane. Sida Decentralised Evaluation 2018:8.
10. *The key role of ICTJ in the designing of Colombia's complex Transitional Justice System.* Final Report. Jocke Nyberg, Heidi Abuchaibe Abuchaibe, Daniela Martínez Pérez. Sida Decentralised Evaluation 2018:9.
11. *Mid Term Review of SymbioCity Kenya.The Sustainable Urban Development Programme in Kenya – 2015–2018.* Final Report. Leif Danielsson, Tim Greenhow, Lucy Njigua. Sida Decentralised Evaluation 2018:10.

12. *Evaluation of AAU's Core Programmes and projects under the Core Programme 2013–2017, with particular focus on Sida's institutional and program support 2013–2017*. Final Report. Adam Pain, Catherine Ngugi, Michael Cross, Jennifer Sesabo. Sida Decentralised Evaluation 2018:11.
13. *Evaluation of CODESRIA's programme cycle "Forty Years of Social Research and Knowledge Production: Consolidating Achievements and Reaching New Frontiers 2012–2016"*. Final Report. Adam Pain, Stephen Webber, Jennifer Sesabo, Jerome Gouzou. Sida Decentralised Evaluation 2018:12.
14. *Evaluation of Swedish International Training Programmes (ITP); Quality Infrastructure – Technical Barriers to Trade (304) and Sanitary/Phytosanitary (305) 2013–2017*. Final Report. Eric Buhl-Nielsen, Mimi Groenbech, Stefano Sedola, Gemunu Wijesena, Jonas Noren. Sida Decentralised Evaluation 2018:13.
15. *Evaluation of Swedish International Training Programmes (ITP); Private Sector Development (Private Sector Growth Strategies and Strategic Business Management) 2011–2016*. Final Report. Sida Decentralised Evaluation 2018:14.
16. *Evaluation of the African Organisation of English-speaking Supreme Audit Institutions (AFROSAI-E)*. Final Report. Greg Moran, Marion Baumgart dos Santos, Kjeld Elkjaer. Sida Decentralised Evaluation 2018:15.
17. *Evaluation of Swedish Civil Society Support in Albania*. Final Report. Sida Decentralised Evaluation 2018:16.
18. *Evaluation of the Pungwe Basin Transboundary Integrated Water Resources Management and Development Programme (PP2)*. Final Report. Svend Erik Sørensen,Troels Kolster, Wellington Dzvairo, Tjitske Leemans, Klas Sandström. Sida Decentralised Evaluation 2018:17.

19. *Evaluation of the Sida supported programme"International Science Programme 2014–2018".* Final Report. Adam Pain, Trish Silkin, Gonçalo Carneiro. Sida Decentralised Evaluation 2018:18.

20. *Mid-term Evaluation of Small-holder Agriculture Reform through Enterprise Development (SHARED) Project, iDE*. Final Report. Eric Buhl-Nielsen, Mimi Groenbech, Stephen Tembo, Emelie Pellby. Sida Decentralised Evaluation 2018:19.

21. *Mid-term Review of Diakonia's Human Rights and Democracy Programme in Cambodia 2017–2019*. Final Report. Annika Nilsson, Kimsrun Chhiv. Sida Decentralised Evaluation 2018:20.

22. *Mid-Term Evaluation of "Regional Economic Integration through the Adoption of Competition and Consumer Policies in the Middle East and North Africa (COMPAL GLOBAL-MENA)"* Final Report. Kimiko Hibri Pedersen, John Lawrance, Richard Moody. Sida Decentralised Evaluation 2018:21.

23. *Mid-term Review of the EMPOWER Private Sector Project.* Final Report. Sarah Gray, Ulf Färnsveden, Art Shala. Sida Decentralised Evaluation 2018:22.

24. *Evaluation of four NGO implemented programmes in Georgia.* Final Report. Pierre Walther, Vera Devine, Nino Partskhaladze, Mikheil Kukava. Sida Decentralised Evaluation 2018:23.

25. *The Evaluation of the Dairy Hub and Dairy Academy Development Project in Bangladesh.* Final Report. Bo Tengnäs, Shibaji Roy. Sida Decentralised Evaluation 2018:24.

26. *Mid-Term Review of the Project 'Promoting Agriculture, Climate and Trade Linkages in the East African Community 2 (PACT EAC 2)'* Final Report. Greg Moran, Flor E Healy, Julia Leiss. Sida Decentralised Evaluation 2018:25.

27. *Evaluation of Health Guarantee to Centenary Rural Development Bank in Uganda.* Final Report. Ananda S. Millard, Ulrika Enemark, Basil Kandyomunda. Sida Decentralised Evaluation 2018:26.

28. *Mid-Term Review of Sida's regional core support (2014–2019) to the Eastern African Grain Council promoting grain trade in the East African region "Strengthening Regional Grain Markets II"* Final Report. George Gray, Sarah Gray. Sida Decentralised Evaluation 2018:27.
29. *Evaluation of the trade policy training centre in Africa (trapca).* Final Report. Jonas Lövkrona, Mark Pearson. Sida Decentralised Evaluation 2018:28.

## 2019

1. *Evaluation of the Sida supported research capacity building programme* "International Foundation for Science". Final Report. Pierre Walther, Bassirou Bonfoh, Thi Van Ha Nguyen. Sida Decentralised Evaluation 2019:1.
2. *Evaluation of the Strategic Plan 2014–2018 of the International Lesbian, Gay, Bisexual, Trans and Intersex Association (ILGA)* Final Report. Jocke Nyberg, Julie Thaarup, Larissa Arroyo Navarrete, Anna Lidström. Sida Decentralised Evaluation 2019:2.
3. *Evaluation of Swedish Leadership for Sustainable Development.* Final Report. Cecilia M Ljungman, Jens Andersson, Jonas Norén, Emelie Pellby. Sida Decentralised Evaluation 2019:3.
4. *Evaluation of the support to UNICEF to strengthen Child Protection in Abkhazia 2015–2018*. Final Report. Sida Decentralised Evaluation 2019:4.
5. *Evaluation of the Global Partnership for Education (GPE) – Literacy and Numeracy Education Support (LANES) program in Tanzania (2014–2018).* Final Report. Criana Connal, Annelie Strath, Khalid Dihenga. Sida Decentralised Evaluation 2019:5.

6. *Evaluation of ILO Road to Jobs (R2J), Afghanistan.* Final Report. Francis Watkins, Kimiko Hibri Pedersen, Naweed Rahmani. Sida Decentralised Evaluation 2019:6.
7. *Mid Term Evaluation of the Sida and USAID Loan Portfolio Co-Guarantee and The Mali Finance for Food Security and Women Entrepreneurs (FFSWE) programme.* Final Report. Frans van Gerwen, Knud-Erik Rosenkrantz, Hammou Haidara, Nadia Masri-Pedersen. Sida Decentralised Evaluation 2019:10.
8. *Evaluation of UN Women Country Programme in Afghanistan.* Final Report. Francis Watkins, Sippi Azarbaijani-Moghaddam. Sida Decentralised Evaluation 2019:14.
9. *Evaluation of the Guttmacher Institute programme on the use of SRHR evidence to inform action (2015–2018).* Final Report. Annica Holmberg, Philip Wambua, Abena Acheampong. Sida Decentralised Evaluation 2019:15.
10. *Mid-term Review of the Strategy for Sweden's Development Cooperation with Ethiopia, 2016–2020.* Final Report. Svend Erik Sørensen, Kristina Mastroianni, Imeru Tamrat, Matilda Svedberg. Sida Decentralised Evaluation 2019:16.
11. *Evaluation of IBON International and the CSO Partnership for Development Effectiveness (CPDE) Project.* Final Report. Annika Nilsson, Catherine Ngugi, Ellen Girerd Barclay, Ica Fernandez, Jonas Norén, Niels Dabelstein. Sida Decentralised Evaluation 2019:17.
12. *Evaluation of ISET Policy Institute Promoting Georgia's Development through Independent Policy Analysis, Trainings and Civil Society Engagement.* Final Report. Pierre Walther, Nino Parthaladze. Sida Decentralised Evaluation 2019:18.

13. *Evaluation of the Capacity Building for Improvement of Land Administration and Procedures in Bosnia and Herzegovina (CILAP project).* Final Report. Svend Erik Sörensen, Bozena Lipej, Catharina Schmitz. Sida Decentralised Evaluation 2019:19.
14. *Review of Scandinavian Consortium Technical Assistance to Strengthening the Institutional Capacity of the Mozambican National Statistical System (2008–2017).* Review Report. Frans van Gerwen, Norah Teopista, Wamayi Madaya. Sida Decentralised Evaluation 2019:20.
15. *Mid-term Evaluation of Women in Financial Inclusion and Capabilities Project in Zambia.* Final Report. Bente Topsøe-Jensen, Bethel Nakaponda, Stephen Tempo, Frida Ericsson. Sida Decentralised Evaluation 2019:21.
16. *Evaluation of Confidence in Georgia.* Final Report. Ian Christoplos, Joseph Salukvadze, Thomas Dubois. Sida Decentralised Evaluation 2019:22.
17. *Evaluation of Sida's Support to the Project "Organic Trade and Value Chain Development in East Africa" 2014–2019.* Final Report. NIRAS Sweden AB. Sida Decentralised Evaluation 2019:23.
18. *Organising parliamentarians against armed violence Evaluation of Parliamentary Forum on Small Arms and Light Weapons (PFSALW) 2014–2018.* Final Report. FCG Swedish Development AB. Sida Decentralised Evaluation 2019:24.
19. *Evaluation of the International Disability Alliance (IDA) 2015–2018.* Final Report. NIRAS Sweden AB. Sida Decentralised Evaluation 2019:25.
20. *Evaluation of the Sida-supported development cooperation with the Kenya Revenue Authority.* Final Report. NIRAS Sweden AB. Sida Decentralised Evaluation 2019:26.

21. *Evaluation of Twaweza Strategic Plan 2015–2018, Programme Support in Tanzania. Final Report.* NIRAS Sweden AB. Sida Decentralised Evaluation 2019:27.
22. *Evaluation of the Zimbabwe Election Support Network Project 2016–2018*. Final Report. NIRAS Sweden AB. Sida Decentralised Evaluation 2019:28.
23. *Midterm review of the Sida core support to KCSF.* Final Report. FCG Swedish Development AB. Sida Decentralised Evaluation 2019:29.
24. *Evaluation of the Improving Court Efficiency and Accountability of Judges and Prosecutors in BIH Phase 2*. Evaluation Report. NIRAS Sweden AB. Sida Decentralised Evaluation 2019:31.
25. *Evaluation of the regional development programme "Towards a Non-Toxic Environment in South-East Asia", Phase II.* Final Report. NIRAS Sweden AB. Sida Decentralised Evaluation 2019:32.
26. *End of Project Evaluation of the BBC Media Action Radio Waves and Tikambe projects in Zambia.* Final Report. NIRAS Sweden AB. Sida Decentralised Evaluation 2019:33.

## Strategic evaluations (6)

1. *Evaluation of Policy Dialogue as an Instrument – the case of Gender Equality*. Dana Peebles, Jonas Lövkrona, Nadine Jub. Sida Evaluation 2015:01.
2. *Support to Capacity Development – Identifying Good Practice in Swedish Development Cooperation*. Sida Evaluation Report for the Joint Scandinavian Evaluation of Support to Capacity Development. Gonçalo Carneiro, Kristina Boman, Birgitte Woel, Arne Nylund. Sida Evaluation 2015:02.

3. *Evaluation of Sida's use of guarantees for market development and poverty reduction*. Hans Slegtenhorst, Mart Nugteren, Alwin de Haas, Rien Strootman, Marie Heydenreich, Paulo Luswata, Nino Serdarevic, Anders Grettve and Bart Schaap, Sida Evaluation 2016:01.
4. *Evaluation of Sida's ITP approach for Capacity Development*, Ingela Ternström, Elizabeth Bamwine, Marina Buch Kristensen, Japhet Makongo, Björn Ternström, Paul Balogun. Sida Evaluation, 2017:01.
5. *Evaluation of Sida's Global Challenge Funds. Lessons From a Decade Long Journey*. David Smith, Matthew Kentridge, Martin Wright, Juliette Seibold, Katharine May and Theo Sands, Sida Evaluation 2018:01.
6. *Evaluation of Sida's Support to Peacebuilding in Conflict and Post Conflict Contexts*. Erik Bryld, Julian Brett, Nadia Masri-Pedersen, Cécile Collin, Sida Evaluation 2019:01. + background reports.

# Annex 4. The effectiveness criterion

One of several questions of interpretation about the effectiveness criterion in aid evaluation concerns its relationship to the two OECD criteria of impact and sustainability. Table A4.1 below sets out the definitions before 2020 (which is the focus of this report).

**Table A4.1: Definitions of three evaluation criteria**

| Effectiveness | Impact | Sustainability |
|---|---|---|
| "The extent to which a development intervention has achieved its objectives, taking their relative importance into account." | "Positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended." | "The continuation or longevity of benefits from a development intervention after the cessation of development assistance." |

Source: OECD/DAC

An examination of these definitions shows that the criteria overlap. My conclusion is that impact and sustainability add unforeseen (positive and negative) results and sustainability when the aid is withdrawn to the foreseen results of the effectiveness criterion. The other dimensions are included in the effectiveness criterion. The time perspective does not always separate them since objectives are naturally formulated in both the long and the short term depending on what is expected to be achieved. In aid, objectives are often (not always) long term

If analyses of effectiveness only examine short-term or medium-term results, it is (as a rule) not an exhaustive or fully reliable analysis of effectiveness that is carried out. The value of such an analysis for decision-makers is limited since it is not obvious that it says anything about the effectiveness of the activities. Nor is it possible, in the next stage, to then make a full assessment of the intervention's efficiency.

Many aid interventions have a long time horizon and naturally, but to different extents, long-term impact objectives. Like impact, the effectiveness criterion requires assessments of causality Analyses of the sustainability of results after aid is terminated require analysis of results (some require an assessment of causality). Reported effectiveness that does not take account of the sustainability result is probably partly of lower value.

Vedung (2012) describes the relation between effectiveness and side effects in the following way.

> "The fact that the side effect model is nonetheless logically based on objectives is seen from the expression side effects. Side effects can only exist in relation to main effects, which are identified by the objectives of the intervention." (Ibid).

That is, impacts are generally expressed in the form of objectives. Over and above this, there are only unforeseen results.

How should these overlaps be viewed? One risk is that they promote an incorrect picture among evaluators and commissioners in which they assume that objectives fulfilled (effectiveness) can be assessed without taking account of causality or making assessments against agreed overall objectives. Alternatively, it is assumed that the effectiveness and impact criteria in the evaluation correspond to genuinely separate analysis steps or that, for example, it is easier to evaluate effectiveness than impacts.

For this study, this discussion means that we start strictly from the definition of effectiveness as it was specified before 2020 (see above). In line with this, however, its focus on interventions having reached their objectives has made it important to look for answers to the fundamental question (of effectiveness) in the reports as a whole and not just in chapter that refer to effectiveness explicitly or in a heading. With overlapping criteria it is therefore also necessary to analyse what is said in assessments of, for example, impacts in the reports so as to be able to assess effectiveness. I have had to go through the evaluations as a whole, covering texts about effectiveness, texts about impacts and the report as a whole in order to form a full picture of what the reports have to say about effectiveness.

At the same time the effectiveness criterion has been reformulated as of 2020.

**Table A4.2: New and former definition of effectiveness**

| Previous effectiveness criterion (used in this report) | New effectiveness criterion (since 2020) |
|---|---|
| "The extent to which a development intervention's objectives were achieved, or is expected to be achieved, taking into account their relative importance" | "The extent to which the intervention achieved, or is expected to achieve, its objectives, and its results, including any differential results across groups". |

Source: OECD/DAC.

The following reasons are given for the change:

> "Effectiveness and impact are complementary criteria focusing on different levels of the results chain. Effectiveness considers the achievement of results relative to an intervention's objectives, namely at the output and outcome level whereas impact focuses on higher-level results, namely what the declared higher-level results are and what contributes to these. In general, intervention managers and evaluations should ensure that a clear distinction is made between the different results levels (i.e. input, output, outcome and impact) and that it is clear which aspects will be evaluated under each criterion."

One difficulty of this solution is that the effectiveness criterion is then not about effectiveness. Effectiveness is assessed *by definition* relative to what the intervention was intended to achieve (which justified financing it) and that expectation is rarely to be found at the output or outcome level in Swedish or international aid projects. Here again, the new effectiveness criterion seems to go against how effectiveness has traditionally been defined, understood and applied in the field of evaluation (Eliasson (2006), Vedung (2012), Faugert and Sandberg (2012). According to OECD DAC, the purpose of the change is to get away from the overlap relative the impact criterion, but there is a risk that the new criterion will, at the same time, make evaluations less relevant to decision-makers through more reporting of short-term results that say little or nothing about the effectiveness of the intervention.

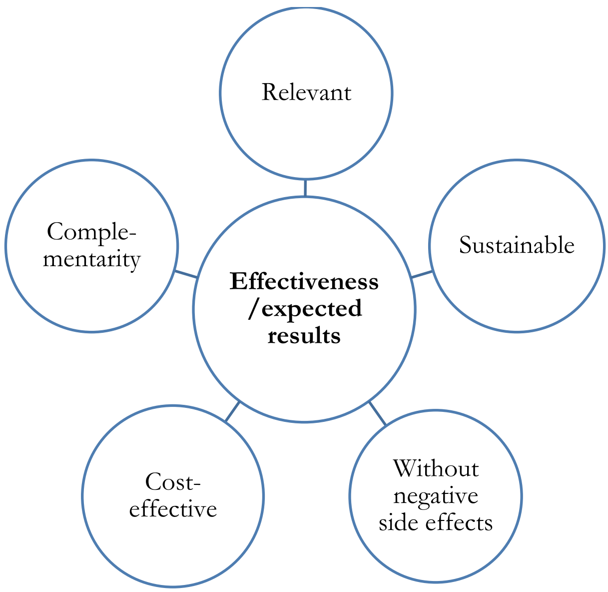
A better alternative for the future might have been to interpret the other criteria as qualifications of the effectiveness analysis/assessment against expected results. One point in separating the OECD/DAC criteria is still valid if the criteria make clear dimensions or qualities that are not missed when the evaluation is initiated.[80]

The OECD/DAC's other criteria are also important. It is a waste of taxpayers' money if aid interventions do not reach their long-term objectives, but it is also a waste if interventions (that reach their objectives) are not relevant to the poor and oppressed (cf. Riddell 2007) or are achieved at a relatively high cost. Taken together, these considerations open the way to an interpretation in which effectiveness is placed at the centre of the assessment instead of being one of five criteria) and is supplemented, for a holistic perspective, by relevance, sustainability, efficiency and the absence

[80] Sida's previous evaluation handbook (Molund and Schill, 2007) took the view that the overlaps are unproblematic since they contribute different dimensions/points of entry that are important in assessments of aid interventions.

of negative side effects.[81] The effectiveness criterion is also of central importance since efficiency and sustainability are (mainly) assessed in the stage after effectiveness In the same way, positive or negative side effects need to be understood relative to ordinary effectiveness (figure A4.1).

**Figure A4.1: The effectiveness criterion is of central importance**




The model links back to the idea that effectiveness and expected results are the core of the evaluation commission (cf. Nachmias quoted in Vedung 2012). But the intervention must *also* be relevant, efficient, sustainable and not lead to harmful side effects in communities and societies where development assistance operates.

[81] Conversely, it would probably not be appropriate to place relevance or coherence at the centre of the analysis since this results in a process focus in evaluation activities.

# Previous EBA reports

2021:1, *Data Science Methods in Development Evaluation: Exploring the Potential*, Gustav Engström and Jonas Nóren

2020:07 *Effects of Swedish and International Democracy Aid*, Miguel Niño-Zarazúa, Rachel M. Gisselquist, Ana Horigoshi, Melissa Samarin and Kunal Sen

2020:06 *Sextortion: Corruption and Gender-Based Violence,* Åsa Eldén, Dolores Calvo, Elin Bjarnegård, Silje Lundgren and Sofia Jonsson

2020:05 *In Proper Organization we Trust – Trust in Interorganizational Aid relations,* Susanna Alexius and Janet Vähämäki

2020:04 *Institution Building in Practice: An Evaluation of Swedish Central Authorities' Reform Cooperation in the Western Balkans*, Richard Allen, Giorgio Ferrari, Krenar Loshi, Númi Östlund and Dejana Razić Ilić

2020:03 *Biståndets förvaltningskostnader För stora? Eller kanske för små*?, Daniel Tarschys

2020:02 *Evaluation of the Swedish Climate Change Initiative, 2009–2012,* Jane Burt, John Colvin, Mehjabeen Abidi Habib, Miriam Kugele, Mutizwa Mukute, Jessica Wilson

2020:01 *Mobilising Private Development Finance: Implications for Overall Aid Allocations*, Polly Meeks, Matthew Gouett and Samantha Attridge

2019:09 *Democracy in African Governance: Seeing and Doing it Differently,* Göran Hydén with assistance from Maria Buch Kristensen

2019:08 *Fishing Aid – Mapping and Synthesising Evidence in Support of SDG 14 Fisheries Targets,* Gonçalo Carneiro, Raphaëlle Bisiaux, Mary Frances Davidson, Tumi Tómasson with Jonas Bjärnstedt

2019:07 *Applying a Masculinities Lens to the Gendered Impacts of Social Safety Nets*, Meagan Dooley, Abby Fried, Ruti Levtov, Kate Doyle, Jeni Klugman and Gary Barker

2019:06 *Joint Nordic Organisational Assessment of the Nordic Development Fund (NDF)*, Stephen Spratt, Eilís Lawlor, Kris Prasada Rao and Mira Berger

2019:05 *Impact of Civil Society Anti-Discrimination Initiatives: A Rapid Review*, Rachel Marcus, Dhruva Mathur and Andrew Shepherd

2019:August *Migration and Development: the Role for Development Aid*, Robert E.B. Lucas (joint with the Migration Studies Delegation, Delmi, published as Delmi Research overview 2019:5)

2019:04 *Building on a Foundation Stone: the Long-Term Impacts of a Local Infrastructure and Governance Program in Cambodia*, Ariel BenYishay, Brad Parks, Rachel Trichler, Christian Baehr, Daniel Aboagye and Punwath Prum

2019:03 *Supporting State Building for Democratisation? A Study of 20 years of Swedish Democracy Aid to Cambodia,* Henny Andersen, Karl-Anders Larsson och Joakim Öjendal

2019:02 *Fit for Fragility? An Exploration of Risk Stakeholders and Systems Inside Sida*, Nilima Gulrajani and Linnea Mills

2019:01 *Skandaler, opinioner och anseende: Biståndet i ett medialiserat samhälle*, Maria Grafström and Karolina Windell

2018:10 *Nation Building in a Fractured Country: An Evaluation of Swedish Cooperation in Economic Development with Bosnia and Herzegovina 1995–2018*, Claes Lindahl, Julie Lindahl, Mikael Söderbäck and Tamara Ivankovic

2018:09 *Underfunded Appeals: Understanding the Consequences, Improving the System,* Sophia Swithern

2018:08 *Seeking Balanced Ownership in Changing Development Cooperation Relationships*, Nils Keizer, Stephan Klingebiel, Charlotte Örnemark, Fabian Scholtes

2018:07 *Putting Priority into Practice: Sida's Implementation of its Plan for Gender Integration*, Elin Bjarnegård, Fredrik Uggla

2018:06 *Swedish Aid in the Era of Shrinking Space – the Case of Turkey*, Åsa Eldén, Paul T. Levin

2018:05 *Who Makes the Decision on Swedish Aid Funding? An Overview*, Expertgruppen för Biståndsanalys

2018:04 *Budget Support, Poverty and Corruption: A Review of the Evidence*, Geske Dijkstra

2018:03 *How predictable is Swedish aid? A study of exchange rate volatility,* Númi Östlund

2018:02 *Building Bridges Between International Humanitarian and Development Responses to Forced Migration*, Alexander Kocks, Ruben Wedel, Hanne Roggemann, Helge Roxin (joint with the German Institute for Development Evaluation, DEval)

2018:01 *DFIs and Development Impact: an evaluation of Swedfund,* Stephen Spratt, Peter O'Flynn, Justin Flynn

2017:12 *Livslängd och livskraft: Vad säger utvärderingar om svenska biståndsinsatsers hållbarhet?* Expertgruppen för biståndsanalys

2017:11 *Sweden's Financing of UN Funds and Programmes: Analyzing the Past, Looking to the Future*, Stephen Browne, Nina Connelly, Thomas G. Weiss

2017:10 *Seven Steps to Evidence-Based Anticorruption: A Roadmap*, Alina Mungiu-Pippidi

2017:09 *Geospatial analysis of aid: A new approach to aid evaluation*, Ann-Sofie Isaksson

2017:08 *Research capacity in the new global development agenda*, Måns Fellesson

2017:07 *Research Aid Revisited – a historically grounded analysis of future prospects and policy options*, David Nilsson, Sverker Sörlin

2017:06 *Confronting the Contradiction – An exploration into the dual purpose of accountability and learning in aid evaluation*, Hilde Reinertsen, Kristian Bjørkdahl, Desmond McNeill

2017:05 *Local peacebuilding – challenges and opportunities*, Joakim Öjendal, Hanna Leonardsson, Martin Lundqvist

2017:04 *Enprocentmålet – en kritisk essä*, Lars Anell

2017:03 *Animal health in development – it's role for poverty reduction and human welfare,* Jonathan Rushton, Arvid Uggla, Ulf Magnusson

2017:02 *Do Anti-Discrimination Measures Reduce Poverty Among Marginalised Social Groups?* Rachel Marcus, Anna Mdee, Ella Page

2017:01 *Making Waves: Implications of the irregular migration and refugee situation on Official Development Assistance spending and practices in Europe*, Anna Knoll, Andrew Sherriff

2016:11 *Revitalising the policy for global development,* Per Molander

2016:10 *Swedish Development Cooperation with Tanzania – Has It Helped the Poor?* Mark McGillivray, David Carpenter, Oliver Morrissey, Julie Thaarup

2016:09 *Exploring Donorship – Internal Factors in Swedish Aid to Uganda,* Stein-Erik Kruse

2016:08, *Sustaining a development policy: results and responsibility for the Swedish policy for global development,* Måns Fellesson, Lisa Román

2016:07 *Towards an Alternative Development Management Paradigm?* Cathy Shutt

2016:06 *Vem beslutar om svenska biståndsmedel? En översikt,* Expertgruppen för biståndsanalys

2016:05 *Pathways to change: Evaluating development interventions with Qualitative Comparative Analysis (QCA),* Barbara Befani

2016:04 *Swedish responsibility and the United Nations Sustainable Development Goals,* Magdalena Bexell, Kristina Jönsson

2016:03 *Capturing complexity and context: evaluating aid to education,* Joel Samoff, Jane Leer, Michelle Reddy

2016:02 *Education in developing countries what policies and programmes affect learning and time in school?* Amy Damon, Paul Glewwe, Suzanne Wisniewski, Bixuan Sun

2016:01 *Support to regional cooperation and integration in Africa – what works and why?* Fredrik Söderbaum, Therese Brolin

2015:09 *In search of double dividends from climate change interventions evidence from forest conservation and household energy transitions,* G. Köhlin, S.K. Pattanayak, E. Sills, E. Mattsson, M. Ostwald, A. Salas, D. Ternald

2015:08 *Business and human rights in development cooperation – has Sweden incorporated the UN guiding principles?* Rasmus Klocker Larsen, Sandra Atler

2015:07 *Making development work: the quality of government approach,* Bo Rothstein and Marcus Tannenberg

2015:06 *Now open for business: joint development initiatives between the private and public sectors in development cooperation,* Sara Johansson de Silva, Ari Kokko and Hanna Norberg

2015:05 *Has Sweden injected realism into public financial management reforms in partner countries?* Matt Andrews

2015:04 *Youth, entrepreneurship and development*, Kjetil Bjorvatn

2015:03 *Concentration difficulties? An analysis of Swedish aid proliferation*, Rune Jansen Hagen

2015:02 *Utvärdering av svenskt bistånd – en kartläggning*, Expertgruppen för biståndsanalys

2015:01 *Rethinking Civil Society and Support for Democracy*, Richard Youngs

2014:05 *Svenskt statligt internationellt bistånd i Sverige: en översikt*, Expertgruppen för biståndsanalys

2014:04 *The African Development Bank: ready to face the challenges of a changing Africa?* Christopher Humphrey

2014:03 *International party assistance – what do we know about the effects?* Lars Svåsand

2014:02 *Sweden´s development assistance for health – policy options to support the global health 2035 goals,* Gavin Yamey, Helen Saxenian, Robert Hecht, Jesper Sundewall and Dean Jamison

2014:01 *Randomized controlled trials: strengths, weaknesses and policy relevance,* Anders Olofsgård

Denna studie utgår från den övergripande frågan om det svenska biståndet når sina mål. I rapporten analyseras bedömningar av måluppfyllelse och effekter i Sidas decentraliserade utvärderingar under åren 2017–2019. I dessa rapporteras över lag god måluppfyllelse. Rapportförfattaren menar dock att analysen av orsakssamband kan stärkas betydligt, samt att övergripande mål med insatser bör bedömas i högre grad. Slutsatser, som självklart inte behöver vara felaktiga, brister annars i tillförlitlighet.

---

This study starts from the overall question of whether Swedish aid reaches its objectives. The report analyses assessments of effectiveness and impact in Sida's decentralised evaluations in 2017–2019. On the whole, they report good effectiveness. However, the author argues that their analysis of causality can be made much stronger, and that overall goals of interventions should be assessed to a greater extent. Otherwise conclusions, which obviously do not need to be wrong, are not sufficiently reliable.

EBA
www.eba.se

*Expertgruppen för biståndsanalys (EBA) är en statlig kommitté som oberoende analyserar och utvärderar svenskt internationellt bistånd.*

*The Expert Group for Aid Studies (EBA) is a government committee with a mandate to independently analyse and evaluate Swedish international development aid.*