

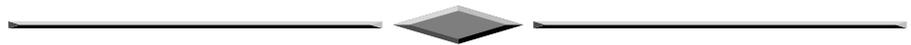


## **ROOM DOCUMENT 6**

# **DAC Network on Development Evaluation**

### **EVALUATING TOTAL ODA EFFECTIVENESS: PROCESS UPDATE AND AMENDED PROPOSAL**

This note has been prepared by Sweden for presentation at the 4th meeting of the DAC Network on Development Evaluation, 30 – 31 March 2006.



**4th meeting  
30 – 31 March 2006**

## 1. Introduction

This note is a follow-up of previous presentations of a proposal for a joint pilot evaluation of the impact of total ODA at country level sponsored by Denmark, the Netherlands, and Sweden (lead agency) with the OECD/DAC secretariat playing a supporting role. It summarises the original proposal, reports on developments since the last OECD/DAC Evaluation Network meeting in June 2005, and presents in brief outline a revised evaluation proposal. As explained below, the total-ODA evaluation project has now been taken to a point where it should be discussed with interested partner countries, if it is to be pursued any further. Remaining questions about the usefulness and feasibility of the proposed evaluation cannot be resolved without partner country participation.

Note that this is a Swedish paper. It has not yet been discussed with the members of the enlarged task force mentioned below. The amended evaluation proposal outlined in the paper is intended to be broadly consistent with the discussions held at the task force meeting in Edinburgh, February 13, 2006, but we do not expect it to match the views of the participants in each and every detail.

## 2. The original proposal

The total ODA evaluation project originated as a response to a call for more comprehensive perspectives in development co-operation evaluations by the DAC Chair, Richard Manning. The need for a broader, more holistic understanding of the effectiveness of ODA has increased with recent changes in the architecture of international development co-operation. While donor objectives are increasingly aligned with the poverty reduction strategies of developing country governments, however, evaluations are still largely concerned with the project and programme level contributions of individual donors. The joint General Budget Support evaluation led by the UK takes a broader perspective, as did recent evaluations of national poverty reduction strategies by the World Bank, IMF and others. Important and useful as they are, however, these evaluations are mainly concerned with implementation processes and intermediary results rather than development impacts, which is the primary concern of the evaluation discussed in this note.

A first proposal of how an evaluation of total ODA impact might be designed and implemented was presented to the OECD/DAC Network on Development Evaluation in November 2004.<sup>1</sup> The proposal can be summarised as follows:

- Focus on *development impact*. The evaluation would be concerned with the impact of ODA on quality of life indicators (longevity, health, food security, education, etc.), economic growth, and other basic development measures rather than with intermediary results.
- Broad coverage. While avoiding the impossible task of assessing the total development impact of total ODA, the evaluation would seek to cover *a reasonably broad range of development impacts*. However, the proposal did not explain in detail what this could mean in practice.
- *Long-term perspective*. Regarded as essential for an assessment of impact.
- *Assessment of ODA in a context of country development*. Starting with country developments and country development priorities was deemed more useful to stakeholders all around than starting with donor inputs and their intended effects. The evaluation would describe how things have changed in relation to selected development indicators over a period of time and then try to understand how ODA has played a part in the process, positively or negatively. To put it even more briefly, it would start with the outcomes and look for their causes rather than the other way around, as in most evaluations of projects and lower-level programmes.
- Tracing the *mechanisms* behind development impacts ('determinants of effectiveness'), not just the impacts themselves. To make it useful for formative purposes, the proposal stressed that the evaluation should try to explain *why* impacts had occurred or failed to occur. It should help us better understand '*what works*' (and what does not work) in the case study countries.
- Focus on the *combined donor support* rather than the contributions of individual actors. It is this aspect of the evaluation more than any other that would make it an evaluation of total ODA. While an evaluation of the entire flow of ODA to a sector or programme could well be described as a

---

<sup>1</sup> Gus Edgren, 2005. Assessing ODA at Country Level: Proposal for a Joint Study. [www.sida.se/total-oda](http://www.sida.se/total-oda)

kind of total ODA evaluation, an evaluation covering only a slice of the ODA allocated to a particular unit could hardly be described as an evaluation of this kind.

- The evaluation should be *a joint undertaking by donors and interested developing countries*. The division of labour with regard to evaluation management should be decided case by case, taking capacity for evaluation management and such matters into account.
- *Independent evaluation team* supervised by representative body of stakeholders. Donors interested in supporting the evaluation would probably want to use it for accountability purposes as well as to promote learning. Partner countries might also want to use it for both types of end. Important therefore that the evaluation team would not be biased in any particular direction.
- *A highly qualified evaluation team*. The team should be able to handle a suitably broad range of instruments for data analysis and impact assessment and it should be well grounded in the country context. Presumably, it would have to include experts from a variety of academic fields. This is a case where a team of highly skilled evaluators would make a big difference.

The response to this proposal was largely positive, but concerns were raised over the scope of the evaluation and its technical feasibility. Would an exercise on such a grand scale not be too costly, and would it not put too much strain on the limited evaluation resources of partner countries? Furthermore, could well-known problems of causal attribution and the construction of credible counterfactuals be solved? Given that development changes in any particular country are the result of a multitude of factors; would it be possible to say anything very precise about the impact of ODA?

A second line of questioning concerned the evaluation process and the nature of the evaluation partnership. Had we not started at the wrong end? Should discussions about design and methodology not be postponed until we had inquired with partner countries whether they would be at all interested in taking part in the evaluation? Who should sit in the driver's seat? What if our developing country partners decided that they would have nothing much to gain from an initiative of this sort?

The conclusions of the discussion was that the technical questions should be further investigated before it could be decided if and how we might try to engage interested partner countries. It was agreed that an evaluation of this kind had to be a co-operative effort and that it should be managed by partner country institutions as far as possible. It was also agreed that the existing task force with Sweden as lead agency should call on suitable external expertise to further investigate the technical feasibility of the evaluation. Sweden offered to organise the assessment process.

### **3. Consultations**

The assessment process consisted of three main steps:

1. Engaging a team of well-established development economists to assess the original proposal in relation to the state of the art in development economics, and, if required, sketch an alternative proposal more in tune with existing constraints on evaluability and current priorities in development research. The assessment team consisted of Prof. Arne Bigsten, University of Gothenburg, Prof. Jan Willem Gunning, Free University of Amsterdam, and Prof. Finn Tarp, University of Copenhagen. The team was asked to take an independent scholarly look at the proposal, not sparing it of criticism.
2. A workshop in Stockholm, November 10-11, 2005, to discuss the assessment by Bigsten et al, including a proposal for an alternative evaluation design. The workshop also discussed some related approaches to macro-level evaluation. Presentations were made by representatives of the Peoples Republic of China, the Republic of South Africa, EC, UNEG, and the World Bank, as well as by Bigsten et al. The workshop participants were development researchers from different academic fields and representatives from Ghana and South Africa as well as Evaluation Network members. The entire OECD/DAC Evaluation Network had been invited.
3. An assessment of a revised proposal from Bigsten et al by an extended OECD/DAC Evaluation task force consisting of Sweden, Denmark, the Netherlands, the OECD/DAC Evaluation Network Secretariat, France, Germany, the UK, EC, and the World Bank. The meeting was held in Edinburgh, February 13, 2006, in connection with a Steering Group Meeting of the GBS Evaluation.

Describing the assessment process in detail is beyond the purview of this note.<sup>2</sup> However, the following are a few of the key conclusions:

- While supporting the proposal in principle, the Bigsten team regarded it as altogether too ambitious. Here they speak for almost everyone who has been involved in the process. If an evaluation of total ODA means looking at everything, as the original proposal may have seemed to suggest, it is just not doable. A strategic selection of modalities, themes, indicators, etc. would be required.
- The proposal from Bigsten et al consisted of two parts. On the one hand it suggested that we should examine the impact of aid on policy reform, or rather that we should take a closer look at the conditions conducive to such impact. Policy reform has been a prominent theme in development economics for a long time, but there are still important questions to be answered. The second component was an experimental study of the aggregate impact of ODA project support at sector level. The first of these two components was presented in the Stockholm workshop and discussed with a great deal of interest, the second was introduced later. The task force meeting in Edinburgh came to the conclusion that while both the proposed studies were interesting in their own right, they did not add up to an evaluation of total ODA. The participants agreed that an evaluation of total ODA impact should focus on development results further down the line. While policy impact might figure importantly among the independent variables of the evaluation; it would not be its main dependent variable, as in the Bigsten proposal.
- The issue of the counterfactual has been much debated during the process. According to the Bigsten group, an impact evaluation is not ‘scientifically defensible’ unless it can establish a credible estimate of the counterfactual, the developments that would have occurred in the absence of the evaluated intervention. Furthermore, according to the team, experimental methods based on randomisation and statistical methods closely modelled on the experimental approach set the standard for what can be regarded as scientifically acceptable procedures for counterfactual analysis. In the Stockholm seminar this view was to some extent disputed. It was argued that historians, political scientists, and others produce credible causal analyses by other means. In evaluations of World Bank country assistance strategies causal analysis is based on a more inclusive concept of ‘plausible association’ that probably better matches work done in some of the social sciences. However, in the end these disagreements seemed somewhat illusory. While Bigsten et al insist on the scientific superiority of randomised and statistically based methods, they freely admit that there are situations where alternative approaches must be used. Indeed, in their own evaluation proposal they recommend the use of such second-best methods. The main lesson for an evaluation of total ODA, it seems, is that it should be methodologically eclectic, while keeping a watchful eye on the limitations of the different methods put to work. The evaluation should also be open to the possibility that there are questions that cannot be answered at all. However, those questions cannot be fully identified in advance. Evaluating total ODA would be a case of learning by doing. Indeed, this is why it is presented as a pilot evaluation.
- The original proposal suggested that partner country development objectives should be used as the main criteria for the evaluation. Thus, the evaluation would in large part be an assessment of the extent to which the total combined ODA has supported the achievement of those objectives. Bigsten et al take issue with this suggestion, pointing out that there are oppressive regimes with bad polices as well as progressive ones promoting development. They also make the obvious point that government policies tend to shift with time and changes of government. However, they do not propose that the original criteria should be replaced by a multitude of shifting donor agency objectives. Their view is rather that the evaluation should be made with reference to “a generally accepted broader set of welfare measures relevant to the poor or the whole of the population.” This seems to be a sensible solution to the question of evaluation criteria. If individual donors want to assess aggregate outcomes against their own particular objectives, they should do so. However, to be useful to partner countries, the evaluation must be based on criteria that closely match the needs and development aspirations of the citizens of those countries.
- One very big question still hangs in the air. This is the question of what a total ODA evaluation would be good for. Who needs it and why? As the DAC Chair originally made his case, comprehensive evaluations on this scale would primarily serve a purpose of accountability vis-à-vis principals and the general public in countries involved in the development co-operation effort as donors or recipients. During the process described in this paper, however, questions about

---

<sup>2</sup> Further information is available at the ODA evaluation web site [www.sida.se/total-oda](http://www.sida.se/total-oda)

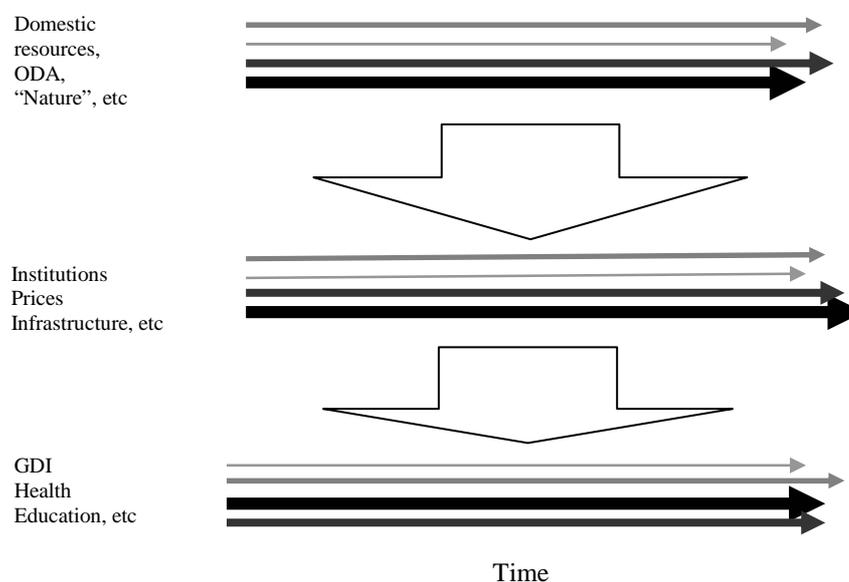
learning and improvement have also come to the fore. Yet, here we encounter a worry often raised with regard to evaluations that take a longer view. What is the point of looking back on modes of operation that we have already put behind us and no longer believe in? Would it not be much more useful to focus on our present agenda? These are questions that everyone must answer for himself. However, in this evaluation proposal it is assumed that the past may not be as different from the present as we often like to think. The proposal also rests on the view that stocktaking is important even when it does not feed directly into our daily operations. If we are at all interested to know what our efforts have amounted to, questions about long-term impact cannot be avoided.

- The question of ownership has come up repeatedly during the process, and at the Stockholm workshop it was one of the main topics for discussion. Echoing earlier debates, several participants felt that interested partner countries should be invited to participate right away, and that postponing their participation would just make things more complicated. Approaching them with an evaluation agenda already fixed would be outright counterproductive. In response to this there was again the view that we should know what we are talking about before we go ahead and formally ask our partners to join us. With the enlarged task force meeting in Edinburgh, however, it seems that we may have completed the process of deliberation started in Paris in 2004. As things now stand, asking interested partner countries to join the discussion would seem to be the logical next thing to do. The main alternative, it seems, would be to put the project in cold store along with other interesting things that we might wish to do in the future.

#### 4. An amended proposal

With the meeting in Edinburgh we may seem to have come full circle. However, this is not really the case. While we have decided to stick to the fundamentals of the original proposal, we have followed the advice of our consultants regarding the scope of the study and the criteria for evaluation. As indicated above the evaluation will focus on the impact of ODA on a restricted set of impact variables, while generally accepted measures of welfare will be used as evaluation criteria. As the main features of the original proposal have been summarised above, what is left to do here is only to briefly clarify a number of points that may still seem obscure.

As we conceive it here, an evaluation of total ODA impact would deal with the three layers of facts distinguished in the figure below.



Starting from the bottom, there is a first series of development impact variables, such as GDI, health, and education. These are the dependent variables on which ODA, in interaction with other factors, is expected to have an impact. As indicated in the figure, the impact of ODA is usually mediated through an intermediary layer of variables, such as institutions, prices, or infrastructure. The top layer in the figure consists of the inputs to the development process. ODA is one of those inputs and so are

domestic resources of various kinds. There is also a broad category of inputs that are exogenous to both donors and recipients. Referred to as “nature” in the figure, this category includes the weather, world market prices, technological change, and other variables not controlled by donors and recipients.

The causal network into which ODA is injected is of course a very complex one. Not only are there ‘vertical’ relations between variables at different levels, as the two arrows pointing downwards suggest. There are also horizontal interactions of many kinds between variables at one and the same level. Clearly, comprehensively tracing the influence of total ODA through this network down to each and every development variable would be much beyond the powers of any single evaluation.

To be manageable the evaluation would have to focus on a smaller subset of the impact variables relevant to the measurement of development. The choice of variables to be covered would presumably be preceded by a review of changes in basic development indicators over time, and a parallel mapping of changing allocations of ODA across the economy and between sectors. Supported by the results of these reviews, the evaluation stakeholders would jointly select a range of variables relevant to their common interest in understanding the mechanisms and impacts of total ODA.

Once the focal development variables have been selected, the main task for the evaluation would be to carefully map changes in those variables over time and to try and understand how those changes have come about. Here the evaluation would not focus on ODA to the exclusion of other contributory factors. Quite the opposite, in the first phase of the analysis one would try to get as full a picture as possible of the entire range of factors responsible for the recorded changes in development outcomes. In terms of the simplified figure above, one would focus on links connecting the bottom level variables with the intermediary variables. Questions about the role of ODA, positive or negative, would appear at a later stage when the analysis turns to the links between the intermediary variables affecting development outcomes and the input factors shaping these intermediary variables.

The extended time frame is crucially important. In many cases, there is a considerable time lag between intervention and intermediary outcome as well as between outcome and impact. While the effects of a price reform, for example, may be felt almost immediately, the impact of an investment in infrastructure, health services or education systems may take many years to materialise. Furthermore, the impacts of interventions are sometimes delayed because of poor sequencing of policy reforms. By adopting a long-term perspective, however, it would be possible to register the cumulative and mutually reinforcing effects of interventions. How far back in time one would want to go would vary from case to case. While the evaluation might be limited to impacts occurring over the last ten to fifteen years, say, the analysis of the processes that have made those impacts possible could in some cases take us further back in time.

Questions about counterfactuals would be important at many points in this evaluation. Often, however, what we would like to know is not really how the actual recorded impact of a particular intervention differs from what would have happened without any intervention at all, but rather how it differs from what would have happened had things been done differently – an alternative type of intervention or a different style of work. Miracle counterfactuals should be avoided, but asking questions about missed opportunities and the like would probably become an important part of the analysis.

It should be noted that the evaluation would not try to assess the impact of ODA by comparing it to what might have happened in its absence. While project impact is often most reliably analysed in a with-and-without framework, such a model of impact assessment obviously makes little sense at the level of total ODA. Furthermore, for the practical purpose of this evaluation the question of what might have happened in the total absence of ODA is hardly even relevant. What we would like to find out is what went well, what went wrong, and – depending on the answers to those questions – what we might do better in the future.

## **5. The evaluation process**

Not much need be said about the evaluation process in this note. The points about ownership, partnership and independence in the original proposal would still seem to be valid. Note however, that there are three organisational levels to be considered, namely those of governance, management, and research. Once the decision to make an evaluation has been made, the questions of ownership and partnership are largely questions about governance. It is essential that both the partners are properly engaged in the design and steering the evaluation process. Ownership questions are also important at the level of management; at least insofar as involvement in management may be a prerequisite for steering and overseeing the evaluation. However, at the management level there may also be other

reasons for the partner country assuming an important role, including those of learning and capacity building. The same is true at the level of research. Considerations of independence would be important throughout. At the governance level independence is mainly a question of setting the rules of the game in such a way that processes of management and research are protected from undue bias and other threats to credibility.

A further point is simply that the evaluation processes should be adapted to the specific opportunities and constraints existing in each case study country. The purpose of the evaluation is not to produce a ponderous piece of research, but to provide stakeholders with an opportunity to take stock and feed useful lessons into current operations and plans for the future. This is an eminently practical purpose. Achieving it may require some early thinking about means and ends.

## **6. Next steps**

As noted above, the preliminary exploratory phase of the evaluation can be regarded as completed. The next step in this evaluation is to decide if we should put it to rest or introduce it to our development country partners as a proposal for joint work.