



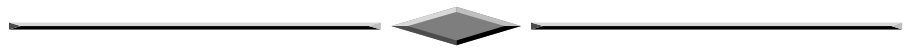
# **ROOM DOCUMENT 11**

## **DAC Network on Development Evaluation**

### **RESULTS EVALUATION AND IMPACT ASSESSMENT**

#### **Item 4: iii**

The attached note has been prepared by the Chair of the DAC Network on Development Evaluation for discussion at the first Network meeting, 15-16 January 2004.



**1st meeting  
15 – 16 January 2004**



## RESULTS EVALUATION AND IMPACT ASSESSMENT

*A draft note for discussion in the meeting of the DAC Network on Development Evaluation, January 15-16, Paris*

Rob D. van den Berg – December 12, 2003

---

The purpose of this note is to enable a discussion on current problems in impact assessment in the light of the increasing emphasis on results.<sup>1</sup> Conceptually, impact assessment faces difficulties in establishing attribution of impact and results to the donor's interventions. The note explores especially the problem of causality, which is central to the question of attribution. It ends with some proposals to reposition evaluation practices in the light of results based management and it explores possibilities for increasing the scope of evaluations into results which matter for the current debate on development.

### *Suggestions for discussion*

1. Because of the expense involved in impact assessment, it is seldom undertaken. Because of the lack of data, it is often restricted in scope. Because of the lack of a counterfactual, it often cannot establish attribution even within its limited scope with any certainty. Nevertheless, evaluation units undertake impact evaluations, with often very interesting results. In view of their limited scope, they offer no more than a glimpse of what development co-operation really contributes to development. The enormous variety of aid in partner countries, sectors, themes and modalities cannot be covered through impact evaluations as we have undertaken them so far. It is clear that we need to reposition impact assessments and evaluations of long-term results of development assistance.
2. This note argues that future evaluations should stress "results" rather than impact as the term which is now central in the debate about development and the management of foreign assistance. However, "results" is a broad category, referring to outputs, outcomes and impacts of policies, programmes and activities. Nevertheless, the focus on results rather than on impact should help in preventing the image that development assistance does not deliver results if impact cannot be shown.
3. The high costs of results evaluation which focus on impact can be shared by undertaking joined evaluations on a country level. The proposal of the DAC Chairman to jointly evaluate the contribution of donors to development in a partner country could be taken up as a welcome initiative in this direction. Other joined work, for example meta-evaluations, in which the DAC Evaluation Network has built up a track record, should be strengthened.
4. The problem of the lack of a counterfactual can be addressed by adopting the historian's definition of causality rather than the one frequently used by statistically inclined social scientists. Development is after all a historical process. Historians never work with a counterfactual, yet are able to shed increasing light on events like the Second World War. Furthermore, good historians use any scientific

---

<sup>1</sup> Impact assessment was discussed previously in the workshop in Washington in October 2002, the Working Party meeting in March 2003 and the 5<sup>th</sup> Evaluation Conference of the World Bank in Washington, July 2003.

discipline that will shed some light on the facts or trends that they study. The care which good historians take to establish facts and trends should be incorporated in our evaluation practices.

5. This does not mean that the methodologies to establish causal linkages through counterfactuals should be disregarded. If and when sufficient data are available, and random experimentation or other forms of establishing a counterfactual are possible, evaluations should use these. This note suggests that this is mainly in innovative programmes with reduced coverage.

6. The general debate about the impact of development assistance refers to causality as something which can no longer be established confidentially when going from project to programme and policy approaches. Instead, it is proposed to establish the “plausibility” of the impact. This note suggests that it would be better to shift to the term “contribution”. Causality in common understanding refers to statements of the sort “A causes B”, which suggest a straightforward direct linkage. The many factors which are involved in causing development to happen bring other analogies to mind – the note discusses machines and biological organisms as more appropriate than the direct linkage model. In these more appropriate analogies “contributions” are made to the greater whole, but it is clear that not one of these contributions is the prime mover or cause of what is happening, and it does not make much sense in these analogies to put a figure on the contribution. Furthermore, we should in our terminology move from “linkages” to necessary but not sufficient “conditions” for change to happen. This should solve the issue of attribution (how much does food contribute to our survival? 100% obviously, but so does oxygen).

7. The evaluation community to a large extent still looks at relevance as a question of coherence or consistency of inputs and activities with policy goals. This means that the focus is very much on ex ante relevance or relevance of planned activities for the policy that is in place for financing these activities. In a time of results based management and budgeting, the question whether or not the **inputs** have been in line with the policies of donors and partner countries is no longer really relevant. The real question is whether the **results** of our actions are in line with the policies and the problems that these policies tried to address. This focus on results instead of inputs does not require a new definition of relevance in the DAC criteria for assessing aid.

### *The problem*

8. Although the importance of impact assessment is recognised by many, there is on the other hand a concern that it is either too difficult or too expensive to really do substantial work on impact assessment. On the other hand, pressure to “show results” is increasing, in the media, politics, boards, parliaments, and the general public. One solution to this – currently being discussed in many donor agencies – is to show results continuously through monitoring systems. Often, these are promoted in systems of “results based management”, “managing for results” or similar public performance management approaches. The role of evaluation in these new systems is often problematic. On the one hand there is recognition that monitoring will tell whether you are doing things right, but not whether you are doing the right thing. Evaluation can, if applied properly, answer the second question. Yet, if evaluation cannot or will not assess impact, or long-term results, will it be able to fulfil this promise? If evaluation limits itself to the study of input, output and outcomes and the process which brought us from one to the other, it can function as “reality check” on monitoring systems, but will have limited value as far as the question “what works and why” is concerned.

### *Impact assessment too expensive?*

9. There are generally speaking two aspects that cause concern on the expense side. The first is the time horizon. Impact assessment faces historical challenges, because the activities that we want to trace the impact from, have long since been finished. Secondly, looking for impact beyond the direct reach of the

activities themselves widens the scope of research considerably. An impact assessment may quickly become very expensive if a very wide scope is chosen and historical reconstruction is part of the research. This often requires extended file and dossier research and analysis of historical data.

10. Even more important for the budget of an evaluation is if crucial sets of data are missing. Many projects and programmes lack baseline surveys, which means that attempts need to be made to reconstruct the situation in which the intervention started. Since impact assessment is concerned with effects beyond the direct reach of the intervention, it will also use basic data available for the region or country. Although there is a growing mountain of data to be found both nationally, in international institutions and on the internet, the use of these data is by no means unproblematic, because of problems of definition, reliability of the data and validity of the methods to gather the data. In many cases additional data will need to be gathered through surveys – increasing the costs of the evaluation.

### ***Impact assessment too difficult?***

11. Conceptually, the difficulties generally have to do with the question of causal linkages and attribution on the one hand and the counterfactual on the other. Although these two issues are to some extent linked (is attribution possible if there is no counterfactual?) they are best treated separately. First of all the question of causal linkages is of crucial importance in the design of the evaluation. What is the impact that you are in fact looking for? What are expected results out of the direct reach of the activities and how were these results supposedly being effectuated? By definition these linkages grow weak beyond the immediate reach of the activities and become more hypothetical in nature. This means that not only are they more difficult to identify, they are also more difficult to establish and to research.

12. The high number of linkages and the hypothetical nature of many of these between outcomes and impact also weaken the attribution of impact to the outcomes of the activities. But if attribution is reduced to the idea that the outcomes of the activities are “just one of many factors contributing to a certain impact”, then the question comes up whether it is all worthwhile. Should we spend a lot of money on an evaluation which will not really establish attributable impact?

13. The question of the counterfactual is above all a methodological one. Can we establish with scientific certainty that a certain impact is indeed caused by outcomes of activities? One scientific methodology that is often proposed to ascertain impact is to identify a counterfactual which shows that without the outcomes of activities, the impact does not take place. Various methodologies can be used to establish a counterfactual, such as random sampling, control groups, double blind research, but also natural experimentation and theoretical counterfactuals through modelling. In all circumstances the search for the counterfactual adds to the expense of the evaluation as well as to the conceptual difficulties in setting up the evaluation. In order to establish a counterfactual, the evaluation will need to cover more: control groups, representative regions in which no interventions took place, additional data – and will become more difficult methodologically.

14. One possible solution to the lack of data, the lack of insights in how society evolves, the need to establish a counterfactual, is to focus on projects and programmes that take place in circumstances where data are available, which undertake well known and well researched interventions. The first objection to this is that this approach has a built in bias: it amounts to looking for evidence (or your lost car keys) under the glare of the street light – it covers some of the street, but not necessarily the part in which you are most interested (or where you actually lost the keys). The second objection is more fundamental: development co-operation by necessity takes place in data-poor circumstances in situations for which we have not found solutions yet. Or to turn it around: if development co-operation takes place in data-rich situations in which solutions to problems are well known the question is justified what it is doing there.

### *The problem of causality*

15. The question of causality is central to establishing impact. The model that evaluations have used in this regard is that of causal linkage, allowing for attribution of observed changes to the intervention. The negative side of this coin is that if there is no straightforward causal linkage, attribution is not possible and impact can not be established. Many evaluations stop short of describing impact precisely because of this. However, the emphasis on causal linkages indeed leads to (public) discussions on whether aid is making a difference or not. If no causal linkage can be established, perhaps we should stop aid – is an opinion which is sometimes voiced in debates on development co-operation. In my opinion, we need to look carefully at what we mean with the term “causal” before we come to such drastic conclusions. Furthermore, if we do not discuss the problem of causality, these drastic conclusions will play an increasing role in the public debate on the merits of development co-operation. This section will discuss various ways in which the term “causality” is used and end with some suggestions for future use of the term and for the use of alternatives.

16. First of all, our use of the concept “cause” is influenced heavily by physics, which is considered to be the most successful natural science. What do physicists mean by causality? A causal relationship in physics is referring to a linkage which has been established both theoretically and empirically. Typically, a causal linkage is not researched by physicists after it has been proven. If a natural law describes that B will happen if you initiate A, then by necessity B happens and a physicist may limit himself or herself to just checking whether A was indeed put in place, because B will then automatically happen. We (and the physicists) know that in an actual situation B may not happen – however, in 99 % of the cases this does not lead the physicist to doubt the causal linkage – he will start to look for intervening factors which prevented the causal linkage from happening. For example: he lets go of an apple in the air, nevertheless it doesn't fall – he starts looking for a string which holds the apple in the air, rather than claiming that the law of gravity suddenly is no longer valid. Instead of looking for a theoretical causality, the physicist will look for the specific causality which applies in this unique case. For us this distinction between general causality (through laws of nature) and specific causality (the apple doesn't fall because it is held up by a string) is important.

17. It is clear that in the social sciences and in development interventions we have no recourse to natural laws like the law of gravity, and thus cannot refer to causality as something that is established through existing scientific theories. In other words, we seem to lack general causality and we need to restrict ourselves to specific causality. However, the social sciences have not given up the idea of more general insights in how society actually functions. In many areas they have discovered more general “laws” which are considered to provide more insight in how dynamics in society are taking place – most of them economic by nature, but some in other areas of the social sciences as well. It is important to note that many of these theories are disciplinary in nature; i.e. deal with an aspect of society rather than with the interaction of aspects. In other words, the social sciences can sometimes help the evaluation community with general theories about specific linkages, but including all linkages and not with interaction of linkages in a complex dynamic situation.

18. Even if there is no general theory about what is happening, scientific methods can help us establish that linkages exist through statistical proof. If data can be gathered on one side of the linkage(s), and data can be gathered on the other side, it is possible to show that there is a statistically significant correlation between these sets of data. This does not provide any insight, and it has to be treated with care. There are many examples of correlations which do not have any meaning. One is: in the Netherlands in the period 1900-1975 there was a very high correlation between the birth rate and the number of storks in the country. Correlations are acceptable if they are related to hypotheses. In general this should not be a problem in evaluations, which are in principle guided by hypotheses.

19. Two important points to note in this regard is that causality established through statistical means does not require a 100 % linkage and that it may not be direct. For example, a 50 % change in data set A may lead to a 25 % change in data set B. The linkage may be indirect, for example through another agent C ( $A \rightarrow C \rightarrow B$ ) or have a common cause ( $C \rightarrow A$  and  $C \rightarrow B$ ).

20. In situations where linkages need to be proved, (social) scientist tend to look for various methods with which to establish beyond scientific doubt that A and B are (statistically) connected. Examples are laboratory research (in which external factors can be eliminated), random experiments, double blind research with control groups, and so on. At the 5th Evaluation Conference at the World Bank the possibility was discussed for policy makers to use random experimentation to see whether policies and programmes are actually producing results. This discussion in itself is very interesting and especially promising when designing, implementing and evaluating innovative and pilot projects and programmes. For macro policies or sector policies these methods are not suitable, because on that level random experimentation is usually not possible. A drawback of these methods is that they often do not provide insights in why the linkage is there. It seems that in this regard they are more monitoring instruments (are we “on track”) than evaluation instruments (are we on the “right” track). These remarks require further discussion, which are outside the scope of this note.

21. In complex situations, there are often too many events to chronicle and too many specific circumstances to highlight for any general causality to be established. There is one field of expertise particularly interested in describing the multitude of specific events and that is history. History is sometimes called an art and sometimes a science – in so far that it is a science, it is multidisciplinary in that it calls on any discipline necessary to explain various aspects of the history that unfolds. In so far that it is an art, it recognises that there is no general theory of society which would be applicable to explain why history did unfold as it did. By necessity, any historical narrative is thus a narrative from the perspective of the historian, and it deals with specific events rather than general laws.

22. Historians are thus interested in the circumstances and events which caused other events to take place as they did. Furthermore, the causality presented in histories has no problem in crossing the border from micro to meso to macro, as is exemplified by this old English saying, which was first coined into writing by the poet George Herbert in 1633:

*For want of a nail the shoe is lost,  
For want of a shoe the horse is lost,  
For want of a horse the rider is lost,  
For want of a rider the battle is lost,  
For want of the battle, the kingdom is lost!<sup>2</sup>*

23. Historical causality should not be derogated. It is first of all the narrative of what happened visibly, on the surface, and secondly the narrative of why it happened as it happened. If the story is convincing, this narrative can be very powerful. Good historical research can be very authoritative, not least because good historians (and historically inclined social scientists) are “triangulators” par excellence, who will try to get evidence from as many sources as possible in order to establish a single historical occurrence. Since these narratives, however broad they become, are concerned with specific historical events, they do not try to establish general truths or insights, and thus are not concerned with providing a counterfactual. The reliability of the narrative is established through “triangulation” of the historical facts, whereas the validity of the story (the interpretation of the facts) is established through reasoning, which can take the form of a judicial argumentation: weighing the pros and cons of a certain interpretation of what

---

<sup>2</sup> George Herbert –*Jacula Prudentum*.

happened and then coming to a considered judgement. Good history will base this as much as possible on a discussion of the facts.

24. It is important to note that to establish historical facts and events, no counterfactual is needed – especially on a societal level. We can speculate how the world would have evolved if Adolf Hitler would not have been born, but this is the realm of (science) fiction rather than the realm of history. Those who claim that evaluations cannot provide any insight in what is happening and why, if no counterfactual is established, tend to overlook history. There is no counterfactual to the Second World War – nevertheless our understanding of what took place and why is steadily increasing.

25. Many evaluations tend to be historical in nature, yet not in methodology, style or care for the establishment of historical facts. They are historical in nature in the sense that they describe what happened and why, rather than carrying out systematic scientific assessments of the linkages between interventions and changes in society. Even though many evaluations are “histories” of the programmes they evaluate, they are not being carried out by historians. Perhaps as a result, facts are sometimes presented as undisputed and triangulation sometimes does not take place at that level (in many discussions in the evaluation community triangulation is limited to the methodological level – using three different methodologies to gather and/or analyse data). If and when evaluations bring us good “histories” of the policies, programmes and projects, there is great value in these, and this great value should be recognised. These histories also provide insights in causalities, even if limited. Or to quote the philosopher Henry David Thoreau: “Some circumstantial evidence is very strong, as when you find a trout in the milk”. Or in other words: a well established series of facts may be stronger than a weak theory of social dynamics.

26. We are thus faced with a wide variety of uses of the terms “causality” and “linkage”. In physics, ideally causality and linkage refer to general laws of the form “for all A, if A then B”. This kind of causality is (almost?) never used in evaluations of development co-operation. No counterfactual is needed – if B does in actual fact not follow A, then some intervening factor must be identified. In many social sciences, causality often refers to a statistical linkage: “for all A, if A then a percentage of change in B can be observed.” For this kind of causality, the establishment of a counterfactual is important. In history, causality and linkage is treated as a specific truth: “for this A under these circumstances, B followed A.” No counterfactual is possible or needed.

### *From causality to plausibility?*

27. In the discussion on results the difficulty of proving causality has been noted time and again. This has led to a move away in terminology from causality to “plausibility”. This means that the linkage no longer has to be proven, it has to be made plausible. This seems to me insufficient. Either the plausibility is made convincing, in which case we are facing the same problems in our evaluation as with causality, or it is unconvincing, in which case we are actually farther away from home and we may be accused of propaganda.

28. The use of the term “plausibility” is especially promoted when a direct causal linkage can no longer be established, for example because too many factors play a role in determining outcomes and results. In the previous section the various ways in which causality and linkage are being used in scientific and historic circles were discussed. Here we are concerned as well with how causality is used in common parlance. The general public tends to use causality as a verb: to “make happen”. Development co-operation is seen as an input to “make happen” development: or “A causes B”. We have to recognise that changes in society cannot be reflected in that kind of statement. There are many examples of changes which are described in other kinds of logic than the “if A, then B” type of syllogism. Two examples come to mind. First of all, a complicated machine such as the motor of a car consists of many parts. Most of these parts are essential to make the motor run, but it is also clear that these parts do not “cause” the motor to run. It is



in fact the design of the motor, or the interaction of the various parts, which causes the motor to run. Each part has its own contribution to make and may be a necessary but not sufficient condition for the functioning of the motor.

29. Another analogy is from biology. A plant grows according to various stimuli and sources of food and light. If certain minerals are abundant in the soil, it will take on certain colours. If the temperature drops, it may temporarily drop its leaves, and so on. Its personal history is a combination of innate possibilities in its genes and the specific circumstances in which it grows. A fertile soil may be a necessary but not sufficient condition for a plant to grow: if the plant does not get any light and is dependent on photosynthesis, it will not successfully grow, however fertile the soil may be.

30. These analogies tell us that in the case of results on the level of society, the public debate should move from the concept of linear causality to the concepts of conditionalities (necessary but not sufficient conditions for changes to occur). Furthermore, it should be made clear that these necessary but not sufficient conditions **contribute to** rather than **cause** the change to take place. This means in my opinion that the development community should move from causality or plausibility to contribution, and from direct linkages to necessary but not sufficient conditions for change.

### *Evaluation in the light of results based management*

31. The search for impact has become more important in recent discussions about development co-operation. Results-based management requires careful thinking about what results are actually being achieved. The Millennium Development Goals have been phrased in results language and are accompanied by a series of potential indicators which will tell whether or not the MDG's have actually been met.

32. Although the rhetoric is very much focused on results, this does not mean that the MDG's or their indicators are sufficiently clear or well-defined to really enable impact evaluations on their achievements. Three problems arise. Firstly, the goals themselves sound simple, yet often conceal problems of definition. For example, what constitutes basic education differs per country – there is no unified definition which will apply to all countries and all circumstances. Secondly, the indicators would in principle “perform” – that is to say, they would indeed indicate whether or not an MDG is on its way of being met, but there are currently no global monitoring systems which provide valid, reliable and uniform data. Thirdly, the contribution of aid in achieving the MDG's is not spelled out and has to be constructed or hypothesised and researched in many cases.

33. Nevertheless, with the change to results based management, the focus could potentially shift in evaluation from an emphasis on effectiveness and efficiency on the level of activities, to the results that these activities are achieving beyond their direct reach. Many evaluation departments are confronted with the wish coming from the political domain that evaluations provide more insight in results of aid.

34. There is of course an overlap in terminology from “effectiveness” as evaluation criterion to “results” as an object of the evaluation. Effectiveness refers to the relationship between outputs and outcomes. An intervention is deemed to be “effective” if the outputs led to the outcomes. An effective intervention is thus an intervention which by definition shows results. However, “effectiveness” as a criterion does not look beyond outcomes into impact. It does not cover results on that level. In the chain of events that is often used in results based approaches – inputs → throughputs → outputs → outcomes → impact – the last three are termed “results”.<sup>3</sup> Evaluations of effectiveness thus provide a narrow perspective on results, but nevertheless a perspective.

---

<sup>3</sup> Glossary of Key Terms in Evaluation and Results Based Management – OECD/DAC – Paris, 2002 [Evaluation and Aid Effectiveness ; 6]

35. The quest for more information about and insights in results is thus a quest to increase the scope of monitoring and evaluation beyond effectiveness into impact, or to phrase it in results based management terms, beyond outcomes into impact. Impact is defined as positive and negative, primary and secondary long term effects produced by a development intervention, directly or indirectly, intended or unintended.<sup>4</sup> The question is then: what changes can be observed to which the development intervention contributed?

36. The evaluation community to a large extent still looks at relevance as a question of coherence or consistency of inputs and activities with policy goals. This means that the focus is very much on ex ante relevance or relevance of planned activities for the policy that is in place for financing these activities. In a time of results based management and budgeting, the question whether or not the **inputs** have been in line with the policies of donors and partner countries is no longer really relevant. The real question is whether the **results** of our actions are in line with the policies and the problems that these policies tried to address. This focus on results instead of inputs does not require a new definition of relevance in the DAC criteria for assessing aid.

---

<sup>4</sup> Ibidem