

## 1.9. Evaluating development assistance: key concepts and issues

Marianna Bacci Tamburlini, Giorgio Tamburlini

### *Introduction*

In this paper, we provide an overview of the concepts used to evaluate development assistance interventions. Our review is the result of a wide, but far from complete, investigation into how agencies (including international, bilateral and Non-Governmental Organizations (NGOs)\* address evaluation, and particularly into their guiding principles. A critical appreciation of how these principles are applied would require more in-depth research that combines information from websites and official documents with other sources such as interviews with agency heads, beneficiaries and aid workers, in order to provide a first-hand account of key issues and obstacles. Our analysis gives priority to strategic choices (why and what to evaluate) rather than to methodological aspects (how and when to evaluate and who evaluates), since the latter can only be meaningful within a clear evaluation philosophy. Finally, we emphasize the intrinsic complexity of the evaluation of development assistance, which stems from the fact that the potential and desired users of development assistance (e.g. donors, agencies, aid workers and beneficiaries) have substantially different interests and points of view.

### *Why evaluate?*

All development assistance evaluation systems agree that the aim of evaluation is to provide information to “do things better”. However, doing things “better” does not necessarily mean doing the “right” things. This is by no means a trivial distinction, since it has clear implications regarding “what” to evaluate. One good definition is provided by the UK’s Department for International Development (DFID): “Evaluation of development assistance describes what has happened and why [...] to capture lessons learned for the future”<sup>1</sup>. The same document highlights that the learning process comprises “openly documenting and disseminating experience for all development partners”, so that “this information can be transformed into knowledge” that will “inform future strategic planning” and involve those in charge of the interventions in discussions about the results achieved by the agency and ultimately to “improve the collective performance of the interna-

\* Our choice of information sources is essentially motivated by the need to include all the different types of agencies (international and national, governmental and non-governmental) as well as by the availability of online material.

tional development community”.

According to this opinion, the “lessons learned” may encompass a specific project, as well as the work of one of more agencies in general, and the validity of the objectives, as well as of the strategies for achieving them. The document also recommends that fostering “a transparent, inquisitive and self-critical organisational culture” be included among the evaluation aims. These principles, recommended by the DFID guide, are shared by the Development Assistance Committee (DAC), which provides guidelines for evaluating development assistance on behalf of the Organization for Economic Cooperation and Development (OECD)<sup>2</sup>. Were these principles implemented, they would enable all development assistance agencies to take a substantial step forward.

A second purpose of evaluation is to be instrumental to improving accountability, which is defined as the process of providing the interested parties, i.e. donors, partners, beneficiaries, etc., with information on how the allocated resources are used. It should be emphasized that the term “beneficiary” includes a range of different actors, including the population groups that are directly targeted by the intervention, civil society groups and local institutions, etc., and that the importance attributed to each beneficiary by the different agencies may vary. For instance, according to the World Bank, accountability mainly refers to the ability of a counterpart government to account for its use of funds. In general, it seems that many agencies are concerned more with accountability to donors than to beneficiaries<sup>\*\*</sup>. Some organizations, such as Save the Children<sup>3</sup>, undergo evaluations of management reliability and efficiency by independent agencies, a practice which is essentially aimed at providing proof of reliability vis-à-vis potential donors, but does not necessarily represent a quality guarantee of strategic contents. Spanish Cooperation also highlights the role of “internal” evaluation in fostering participation, guaranteeing communication flows and supervising process quality<sup>4</sup>. As we mentioned earlier, far more in-depth analysis would be required to assess the extent to which evaluation principles defined by programmatic documents are actually implemented and lessons learned have an impact. Some agencies, such as the World Bank and the European Commission, are transparent in that they make available the full set of evaluations for their projects, but rarely provide information about the impact of the evaluation itself<sup>5, 6</sup>. Ultimately though, changes of strategy are more the consequence of political shifts, both inside and outside the agency, than the consequences of critical appraisals of the interventions. Alongside the declared and ‘politically correct’ purposes of evaluation, there is a range of possible distortions, which have been pointed out by the Swedish International Development Agency (SIDA). These include “ritual” use of evaluation with no real impact, its use for political legitimization of specific approaches, and tactical use in order to promote an agency or an intervention philosophy<sup>7</sup>.

<sup>\*\*</sup> For a more complete insight into this aspect, see “*From Paris 2005 to Accra 2008: will aid become more accountable and effective? Draft for discussion at regional consultations, September 2007*” developed by the International Civil Society Steering Group for the Accra High Level Forum ([www.betteraid.org](http://www.betteraid.org)). We include a key quote from this document “the monitoring of recipient governments has been in large part a review of compliance with norms and standards [...] which are, in many cases, defined by donors”.

### *What to evaluate?*

Most agencies agree with the criteria proposed by the DAC to evaluate development assistance projects, i.e. relevance, effectiveness, efficiency, impact and sustainability<sup>8</sup>. More recently, the notions of harmonization and empowerment have become increasingly popular. Let us have a quick look at these concepts using their DAC definitions.

**Relevance** is the extent to which intervention objectives conform to beneficiary needs, global priorities and donor policies. It includes a combination of compliance with the specific project objectives and/or the organization's general mission, and appropriateness to context. Some confusion may derive from the fact that agencies sometimes review their general objectives before the project is over.

**Effectiveness** is clearly defined as the extent to which objectives have been achieved, or it can be plausibly presumed they can be achieved, also taking their relative importance into account. Its measurement, however, may encounter a host of difficulties, both practical (e.g. availability and reliability of data sources) and methodological (e.g. distinguishing effects attributed to intervention from effects due to external factors).

Little needs to be said about **efficiency**, which measures how cost-effective the conversion of resources into results is, other than it is seldom properly assessed.

A crucial criterion is **impact**. It consists of "primary and secondary, medium- and long-term effects produced intentionally or unintentionally by aid intervention", and, most importantly, may include both positive and negative effects. An impact evaluation examines "changes in the well-being of individuals [...] that can be attributed to a particular project, program or policy"<sup>9</sup>. Evaluating impact means taking into consideration all effects produced directly or indirectly by the intervention. These effects may include changes not explicitly pursued by the intervention, which may, as a consequence, also be undesired. This last aspect should be highlighted as it is almost always overlooked when evaluating development assistance. SIDA lists potential "negative" or "unintended" results for a project<sup>10</sup>, including:

- substitution or displacement: positive effects are achieved for a specific group to the detriment of other groups or at the expense of more important objectives;
- targeting error: interventions ultimately fall short of reaching the target population. For instance, the benefit incidence indicator is used to measure the proportion of benefits reaching the poorest 20% of the population<sup>11</sup>. The surprisingly low level of this indicator in many health projects, even those specifically targeting the poor, confirms the validity of the inverse care law ("The availability of good medical care tends to vary inversely with the need for the population served")<sup>12</sup> and therefore the importance of using this and similar indicators.
- recoil effects: negative consequences such as critically overburdening the state apparatus of a specific sector;
- fungibility: financial resources provided by donors to governments/communities for use in specific sectors (e.g. health) that allow recipient authorities to shift resources to other sectors (e.g. buying weapons)

- perverse effects: effects that are contrary to the aim of the intervention, e.g. increased dependence on external assistance as a consequence of projects aimed at increasing local autonomy.

The overall impact of a project should be at the top of the list of aspects to evaluate. It makes no sense to have a programme that reaches its specific objectives, but produces no overall positive impact, or even a negative one. A problem is that in many cases overall impact, as well as some specific effects, can only be measured in the long-term, therefore it becomes necessary to rely on plausible evidence that implementing intervention  $x$  may produce effects  $y$  and  $z$  in the population.

**Sustainability** is the probability that the interventions included in a specific project and therefore the expected benefits can be sustained once the external assistance is over. The concept is as frequently mentioned as often unachieved, one reason being the volatile nature of bilateral and multilateral development programmes. Ensuring sustainability starts with long-term planning and continuity.

**Harmonization** is the effort of making objectives, indicators and strategies increasingly consistent across agencies. It is a relatively new concept. The international community has recently made clear commitments to comply with this concept, which was officially confirmed in the Paris Declaration on Aid Effectiveness<sup>13</sup>. Harmonizing policies seems, in principle, a logical move towards improved effectiveness. However, it has been argued that “aid creates power relationships between donors, governments and citizens” and that “by crowding together and aligning their policies, donors increase their already significant power over aid recipients, which allows them to keep imposing their priorities and concerns”<sup>14</sup>. Furthermore, if harmonization is heading in the wrong direction, it will generalize the consequent negative effects. Perhaps the more operational concept of “coordination” would be preferable to “harmonization”. Once again, evaluating the impact of harmonized policies would be crucial.

Finally, **empowerment** has also been proposed as a criterion for evaluating development assistance. According to the World Bank, empowerment “is the process of increasing the capacity of individuals or groups to make choices and to transform those choices into desired actions and outcomes”<sup>15</sup>. Pursuing empowerment has become mandatory in development assistance programmes, but this objective is often defined in an exceedingly generic way. This increases the difficulty to evaluate it unless specific indicators are identified: for instance, a recent micro-credit project in South Africa chose the reduction of intimate partner violence as an indicator of women’s empowerment<sup>16</sup>.

### *How to evaluate and when*

We will not cover in detail the whole range of components an evaluation should include (such as the indicators needed to evaluate whether and to what extent objectives have been achieved, or the resources needed for the evaluation and the work plan). Similarly, we will not look into problems regarding the availability and quality of data sources and the choice of analytical tools, such as the statistics used

to evaluate health interventions. Rather, we deem it important to underline the difference between monitoring and evaluation. Evaluation may include monitoring, i.e. a continuous process of collecting information about progress in interventions and use of resources, but implies a broader analysis of results and impact, seeking the reasons, both inside and outside the project, why things went in a certain direction. The core of evaluation lies in asking the right questions, including uncomfortable ones, and finding the most appropriate indicators. The boxes below offer some examples of a hypothetical project on maternal health in which the DAC criteria have been applied in order to identify key questions and relevant indicators.

#### **Relevance**

Questions: Are project objectives consistent with beneficiaries' needs and with the priorities identified by situation analysis regarding pregnancy and delivery? Are planned activities appropriate to reducing maternal mortality and pregnancy complications and are they compatible with the local health system and with ongoing programmes addressing reproductive health? Does the project take the overall social context into account? For instance, if support is to be given to a hospital so that the management of complicated pregnancies can be improved, have the issues of access and transport been looked at? If midwives are to be trained for peripheral health centres, have the issues of their accommodation and transport been looked at?

Examples of possible qualitative and semi-quantitative indicators: compliance of project objectives with local health plan priorities; adherence of planned interventions with internationally recommended effective interventions.

#### **Efficiency**

Questions: Has an explicit limit been set on the proportion of expenses for administration and logistics within the overall budget? Have project resources been adequately allocated to its various components (human resources, supplies, drugs, etc...) and have efforts been made to avoid wastage and use exclusively essential drugs and appropriate technologies? Has there been an effort to minimize costs by coordinating the project with the local system and other projects, such as the pooled procurement of drugs and supplies?

Examples of possible quantitative indicators: cost per service unit, such as per antenatal visit or delivery, or per outcome unit, such as per each avoided maternal death or complication.

#### **Effectiveness**

Questions: Has there been an improvement in any of the process or outcome indicators (i.e. antenatal visits, births attended by skilled personnel, maternal deaths, specific complications, or newborn deaths) compared to reliable figures for the pre-intervention phase or districts/areas with similar characteristics, which were not included in the project?

Examples of possible quantitative indicators: reduction of maternal mortality rate, or of specific obstetric complications (comparison before/after intervention or between districts included/not included in the intervention).

**Impact**

Questions: Have there been any unpredicted effects that can be attributed to the intervention or some of its components? For instance, has there been a drain of skilled workers or any other resource from other health programs or services? Has improved coverage of antenatal and perinatal care produced positive effects on information among women about other health issues, such as family planning or nutrition? Has there been a progressive redistribution (towards greater equality) or regressive redistribution (towards greater inequality) among population groups in terms of access to care, or generally to disposable income? Can any other change in the local community be attributed to some extent to the project?

Examples of possible qualitative or semi-quantitative indicators: improvement or worsening of performance in health programs/services due to resource diversion caused by the project. Change in socially relevant indicators, such as birth-spacing and maternal and child nutritional status.

**Sustainability**

Questions: Will local institutions be able to maintain the activities implemented by the project, both organizationally and financially in the medium-to-long term? For example, are there plans for a capacity building effort in the administrative management of maternities? Are there plans for hospital managers and local administrators to take responsibility for crucial elements such as access to transport and recruitment of skilled health workers?

Examples of possible qualitative or semi-quantitative indicators: existence of a plan to ensure that current activities introduced by the project will be maintained; measurable increase in the management responsibilities of hospital administrators and skills of health professionals.

The answer to the above questions and the calculation of indicators may require the use of information sources other than those routinely used for monitoring purposes, for example, interviews with donors, project leaders and, most importantly, with beneficiaries. With respect to the key issue of impact evaluation, a World Bank document makes some interesting remarks: "To ensure methodological rigor, an impact evaluation must estimate the counterfactual, that is, what would have happened had the project never taken place"<sup>17</sup>. Counterfactual analysis, if rigorously applied, usually requires complex data analysis. However, such an exercise could be highly informative even if carried out "without numbers" and were it based on qualitative information and discussion among key stakeholders.

Although there is clearly a limit on how much can be spent on evaluation, a proper evaluation plan increases the likelihood that the resources will be spent appropriately and that future projects will benefit from the lessons learnt. The evaluation plan should be ready well ahead of the actual implementation phase, though flexible enough to adapt to any changes in the project. Interim evaluative reviews should ensure that the necessary corrections are made along the way. Furthermore, an evaluation plan must be sufficiently transparent to all stakeholders, participatory, keeping in mind that the main objective is learning lessons and not finding culprits for drawbacks.

### *Who evaluates*

The main aid agencies have a specific evaluation body: for example, DFID established the New Independent Advisory Committee on Development Impact<sup>18</sup>; the European Union has its own Evaluation Unit<sup>19\*\*\*</sup>. DFID identifies a series of partners who are interested in the evaluation: donors, beneficiaries, other users and interested parties, as well as representatives of the institutions involved, policy-makers, planners, consultants and aid workers<sup>20</sup>, all of which, depending on the project's characteristics, should be involved in the evaluation process. One of the key issues regards the use of external evaluators, who are indispensable for ensuring independent assessment. Evaluators should be given precise terms of reference (the World Bank provides examples of how to develop them<sup>21</sup>) which usually include consultation with project workers in the field, to ensure a thorough insight into the reasons for successes and failures. DAC developed a website ([www.dac-evaluations-cad.org](http://www.dac-evaluations-cad.org)) that provides summaries, and often the complete text, of evaluations conducted by the various aid agencies, classified by agency, country, sector, evaluation type, and key words. The material is made available by the DAC and the website is run by the Canadian International Development Agency (CIDA).

### *Use of evaluation*

A critical issue is how the evaluation is used. Much too often, evaluations are not used to their full potential and their results are not made available to all stakeholders, including beneficiaries, thus thwarting the key formative intent of evaluation. A senior evaluator at the World Bank lists the main reasons why evaluation results may be inadequately used<sup>22</sup>:

- beneficiaries, or other project partners, do not understand the relevance of evaluation questions/objectives/methods because they have not been sufficiently involved;
- results are not available early enough to benefit the project;
- results do not answer the main questions regarding impact, effectiveness etc.;
- methodology is weak therefore conclusions appear questionable;
- lack of evaluation culture among beneficiaries;
- communication of results is poor or insufficient.

All of the above should be considered while the evaluation is being planned. If the evaluation is conceived as continuous and inclusive process, the risk of its results not being used is certainly minimized.

\*\*\*The European Commission plans its evaluations by issue and geographical area; each year evaluations are conducted in different areas.

## Conclusions

What we have discussed so far should lead to a series of recommendations:

- development assistance agencies should adopt evaluation guidelines. Such guidelines should be the result of internal discussion, but also be consistent with a common internationally acknowledged framework, such as the DAC criteria;
- the evaluation plan should be designed with the aim of providing information for “doing things well” as well as “doing the right things”, and should therefore include the relevant questions and indicators;
- the evaluation plan should not be restricted to the project’s specific objectives but include an evaluation of the broader impact of the intervention;
- the evaluation plan should include interim evaluative reviews to be carried out by project workers at different levels and by the project stakeholders in order to foster communication and joint reviews;
- the evaluations of large projects and programmes should make use of external evaluators and independent data sources;
- the evaluation plan should always include the point of view of beneficiaries, as well as of local authorities, institutions and representatives of the communities.

The complex nature of the issue should not discourage us from critically evaluating development assistance work and how it is implemented. As we have argued, evaluation is all too often neglected, or inappropriately carried out. There is clearly the need to build awareness about the importance of evaluation and promote a stronger attitude towards critical appraisal among the aid community. Aid work may have its own intrinsic value, but aid actors should not be exempted from considering what impact is produced upon the people and situations they are supposed to support. The costs of comprehensive, rigorous evaluation may be affordable only by large organizations, but meaningful evaluations can also be carried out with limited resources.

## References

- <sup>1</sup> DFID, *Guidance on evaluation and review for DFID Staff*, Evaluation Department, July 2005; [www.dfid.gov.uk/aboutdfid/performance/files/guidance-evaluation.pdf](http://www.dfid.gov.uk/aboutdfid/performance/files/guidance-evaluation.pdf)
- <sup>2</sup> OECD/DAC, Development Assistance Committee, Development Co-operation Directorate of the Organization for Economic Co-operation and Development, *Glossary of main terms used in results-based evaluation and management*; [www.oecd.org/dataoecd/14/31/17484948.pdf](http://www.oecd.org/dataoecd/14/31/17484948.pdf)
- <sup>3</sup> Save the Children. Results and evaluations; <http://www.savethechildren.org/about/ratings-evaluations.html>
- <sup>4</sup> Manual de Evaluación de la Cooperación Española, *Aprender para Mejorar*, Ministerio de Asuntos Exteriores y de Cooperación, 2007.
- <sup>5</sup> [www.worldbank.org](http://www.worldbank.org)
- <sup>6</sup> <http://ec.europa.eu/europeaid/how/evaluation/work>
- <sup>7</sup> [www.sida.se](http://www.sida.se); *Looking back moving forward*, SIDA, 2007.
- <sup>8</sup> OECD/DAC, See above.
- <sup>9</sup> <http://web.worldbank.org/wbsite/external/topics/extpoverty/extispma/0,,menuPK:384336~pagePK:149018~piPK:149093~theSitePK:384329,00.html>
- <sup>10</sup> SIDA 2007, See above.

<sup>11</sup> J. Baker, *Directions in development: evaluating the impact of development projects on poverty. A handbook for practitioners*, The World Bank, Washington DC 2000.

<sup>12</sup> J. Tudor Hart, *The inverse care law*, «Lancet» 1971, 1, pp. 405-412.

<sup>13</sup> OECD/DAC, *Paris Declaration on Aid Effectiveness*. High Level Forum, Paris, February 28 – March 2, 2005; [http://www.oecd.org/document/15/0,2340,en\\_2649\\_3236398\\_35401554\\_1\\_1\\_1\\_1,00.html](http://www.oecd.org/document/15/0,2340,en_2649_3236398_35401554_1_1_1_1,00.html)

<sup>14</sup> International Civil Society Steering Group. From Paris 2005 to Accra 2008: will aid be more accountable and effective?; [http://betteraid.org/index.php?option=com\\_content&task=view&id=88&Itemid=26](http://betteraid.org/index.php?option=com_content&task=view&id=88&Itemid=26)

<sup>15</sup> [www.worldbank.org](http://www.worldbank.org)

<sup>16</sup> J.C. Kirr, C.H. Watts, G.R. Hargreaves *et al.*, *Understanding the impact of a microfinance-based intervention on women's empowerment and the reduction of intimate partner violence in South Africa*, «Am J Public Health» 2007, 97, pp. 1794-1802.

<sup>17</sup> J. Baker, See above.

<sup>18</sup> [www.dfid.gov.uk](http://www.dfid.gov.uk)

<sup>19</sup> <http://ec.europa.eu/europeaid/evaluation/index.htm>

<sup>20</sup> DFID, See above.

<sup>21</sup> J. Baker, see above.

<sup>22</sup> M. Bamberger, *Enhancing the utilization of evaluations for evidence-based policy making*, in M. Segone (ed.), *Bridging the gap: the role of monitoring and evaluation in evidence-based policy making*, UNICEF Regional Office CEE/CIS, 2008.