Introduction

Over the last 15–20 years, governments and other (public sector) organizations have been paying much more attention to evaluation. It has become a growth industry in which systems of evaluation exist, with their methodologies, organizational infrastructures, textbooks, and professional societies (Leeuw and Furubo, 2008).

In the development world, the growth of monitoring and evaluation (M&E) in particular has been acknowledged as crucial. Kusek and Rist (2004) have articulated its underlying philosophy. M&E stimulates capacity development within countries and organizations to do their "own" evaluations and to produce their "own" performance data. M&E is not focused on *one type of evaluation,* but concerns all of them, including, for example, ex ante studies, rapid appraisals, process evaluations, cost-benefit analyses, and impact evaluations.

Part of the philosophy of evaluation and therefore also M&E is *to put questions first*. Different questions raise a need for different approaches. If the question an evaluator is confronted with is directed toward understanding what a program or policy is about, what the underlying theory of change or logic is, and what the risk factors are when implementing the program, an evaluability assessment or an ex ante evaluation will be an appropriate route to follow. If the question is focused on the implementation of the program or policy or on the role agencies play, then an implementation analysis or a review of the performance of agencies can be appropriate. This can include an audit or inspection. However, if the question is about whether and to what extent the policy intervention made a significant difference (compared with the status quo, compared with other factors and interventions and with or without side effects), then an impact evaluation is the appropriate answer. This Guidance document looks at the latter type of question and corresponding evaluative inquiry, *impact evaluation*. This document discusses questions of what impact evaluation is about, when it is appropriate, and how to do it.

The Network of Networks for Impact Evaluation (NONIE) was established in 2006 to foster more and better impact evaluations by its membership. NONIE uses the definition of the Organisation for Economic Co-operation and Development's Development Assistance Committee (DAC), defining impacts as "[p]ositive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended" (OECD-DAC, 2002: 24).

The impact evaluations that NONIE pursues are expected to reinforce and complement the broader evaluation work by NONIE members. The DAC definition refers to the *"effects produced by,"* stressing the attribution aspect. This implies an approach to impact evaluation that is about attributing impacts rather than assessing what happened. In most contexts, adequate empirical knowledge about the effects produced by an intervention requires at least an accurate estimate of what *would have occurred* in the absence of the intervention and a comparison with *what has occurred* with the intervention implemented.

Following this line of argument, this document subscribes to a somewhat more comprehensive view on impact than the DAC definition does.

Much of the work on impact evaluation that stresses the attribution problem is in fact about attributing short- and medium-term outcomes (to interventions). In practice, this type of attribution analysis is also referred to as impact evaluation, although (in a strict sense) not within the scope of the DAC evaluation. This document includes a discussion on the latter type of analysis as well as the more long-term effects emphasized in the DAC definition (for further discussion of these issues, see White, 2009).

The purpose of NONIE is to promote more and better impact evaluations among its members. Issues relating to evaluations in general are more effectively dealt with within the parent networks and are thus not the primary focus of NONIE. NONIE will focus on sharing methods and learning by doing to promote the practice of impact evaluation. This Guidance document was developed to support those purposes.

The Guidance document was written by and represents the views of the authors, Frans Leeuw and Jos Vaessen, who were commissioned by NONIE. In writing the document, the authors included previous work by NONIE members and took account of their comments in finalizing the document. Given the fact that perspectives on the definition, scope, and appropriate methods of impact evaluation differ widely among practitioners and other stakeholders, the document should not be taken to represent the agreed positions of all of the individual NONIE members. The current Guidance document, highlighting key conceptual and methodological issues in impact evaluation, provides ample coverage of such topics as delimitation, intervention theory, attribution, and combining methods in impact evaluation. It also presents an introduction to such topics as participatory approaches to impact evaluation and assessing impact for complex interventions. These and other topics, such as the evaluation of new aid modalities and country perspectives to impact evaluation, should be developed further in the future.

Impact evaluation in development assistance has received considerable attention over the last few years. The major reason is that many outside of development agencies believe that achievement of results has been poor, or at best not convincingly established. Many development interventions appear to leave no trace of sustained positive change after they have been terminated, and it is hard to determine the extent to which interventions are making a difference. However, the development world is not "alone" in attaching increasing importance to impact evaluations. In fields such as crime and justice, education, and social welfare, impact evaluations have over the last decade become more and more important.[1] Evidence-based (sometimes "evidence-informed") policies are high on the (political) agenda, and some even refer to the "Evidence Movement" (Rieper et al., 2009). This includes the development of *knowledge repositories,* where results of impact evaluations are summarized. In some fields such as criminology and in some professional associations such as the Campbell Collaboration, methodological standards and scales are used to *grade* impact evaluations,[2] although not without discussion (Leeuw, 2005; Worral, 2002, 2007).

Important reasons for doing impact evaluations are the following:

- Impact evaluations provide evidence on "what works and what doesn't" (under what circumstances) and how large the impact is. As the Independent Evaluation Group (IEG) of the World Bank (IEG, 2005) puts it: measuring outcomes and impacts of an activity and distinguishing these from the influence of other, external factors is one of the rationales behind impact evaluation.
- Measuring impacts and relating the changes in dependent variables to development policies and programs is not something that can be done "from an armchair." Impact evaluation is *the* instrument for these tasks.
- Impact evaluation can gather evidence on the sustainability of effects of interventions.
- Impact evaluations produce information that is relevant from an accountability perspective; they disclose knowledge about the (societal) effects of programs that can be linked to the (fi-

nancial) resources used to reach these effects.
- Individual and organizational learning can be stimulated by doing impact evaluations. This is true for organizations in developing countries but also for donor organizations. Informing decision makers on whether to expand, modify, or eliminate projects, programs, and policies is linked to this point, as is IEG's (2005) argument that impact evaluations enable sponsors, partners, and recipients to compare the effectiveness of alternative interventions.

The authors of this Guidance document believe that the ultimate reason for promoting impact evaluations is to learn about "what works and what doesn't and why" and thus to contribute to the effectiveness of (future) development interventions. In addition to this fundamental motive, impact evaluations have a key role to play in the international drive for better evidence on results and development effectiveness. They are particularly well suited to answering important questions about whether development interventions made a difference (and how cost-effective they were). Well-designed impact evaluations also shed light on why an intervention did or did not work, which can vary across time and space.

Decision makers need better evidence on impact and its causes to ensure that resources are allocated where they can have most impact and to maintain future public funding for international development. The pressures for this are already strong and will increase as resources are scaled up for international development. Without such evidence there is a risk of the case for aid and future funding sources being undermined.

Using the word "effects" and "effectiveness" implies that the changes in the "dependent variable[s]" that are measured within the context of an impact evaluation are *caused* by the intervention under study. The concept of "goal achievement" is used when causality is *not* necessarily present. Goals can also be achieved *independent* of the intervention. Changes in financial or economic situations in the world of health and agriculture or in other social conditions can help realize goal achievement, even in a situation where the "believed-to-be-effective" intervention under review is *not* working.

The question of whether impact evaluation should always attempt to measure all possible impacts is not easy to answer. Impact evaluation involves finding the appropriate balance between the desire to understand and measure the full range of effects in the most rigorous manner possible and the practical need to delimit and prioritize on the basis of interests of stakeholders as well as resource constraints.

Key issues addressed in this document

The guidance is structured around nine key issues in impact evaluation:

1. Identify the (type and scope of the) intervention.
2. Agree on what is valued.
3. Carefully articulate the theories linking interventions to outcomes.
4. Address the attribution problem.
5. Use a mixed-methods approach: the logic of the comparative advantages of methods.
6. Build on existing knowledge relevant to the impact of interventions.
7. Determine if an impact evaluation is feasible and worth the cost.
8. Start collecting the data early.
9. Front-end planning is important.

The discussion of these nine issues constitutes the structure of this Guidance document. The first part, comprising the first six issues, deals with methodological and conceptual issues in impact evaluation and constitutes the core of the document. In addition, a shorter second part focuses on managing impact evaluation and addresses aspects of evaluability, benefits, and costs of impact evaluation and planning.

There is no universally accepted definition of "rigorous" impact evaluation. There are some who equate rigorous impact evaluation with particular methods and designs. Given the diversity in

thinking and practice on the topic and the variety in terms of interventions and contexts in which impact evaluation is being applied, the writing of this document has been guided by three basic premises:

- *No single method is best* for addressing the variety of questions and aspects that might be part of impact evaluations.
- However, depending on the specific questions or objectives of a given impact evaluation, some methods have a *comparative advantage* over others in analyzing a particular question or objective.
- Particular methods or perspectives *complement each other* in providing a more complete "picture" of impact.

Moreover, in our view, rigorous impact evaluation is more than methodological design. Rigorous impact evaluation requires addressing the issues described above in an appropriate manner, especially the core methodological and conceptual issues described in Part I.