

15. Impact Evaluation

Principles and Norms

DEFINITION

15.1 *Impact evaluation* is one of a range of evaluations that may be applied to GRPPs at any given time, but usually after the program has evolved to a steady state. It is commonly defined as the systematic assessment of the effects – positive or negative, intended or unintended – of one or more development interventions on the final welfare outcomes of the affected individuals, households, and communities, and the extent to which these outcomes can be attributed to the development intervention(s). In its most rigorous form, an impact evaluation compares the welfare outcomes of the intervention(s) during the period being evaluated with an explicit *counterfactual* – the hypothetical situation that would have prevailed in the absence of the intervention(s). Different approaches to impact evaluation include quantitative impact evaluation, participatory impact evaluation, and theory-based (program logic) approaches. Good impact evaluations will combine all three approaches. (See standards below.)

Draws on materials presented at the International Workshop on Impact Evaluation for Development, November 15, 2006

NEED FOR IMPACT EVALUATION

15.2 In spite of the increased focus on achieving final development outcomes such as those in the Millennium Development Goals, credible impact evaluation studies, which provide scientific evidence of causal links between ongoing development interventions and final welfare outcomes, are fewer than would be desirable to help guide priority setting in development aid. This is as true for GRPPs as for other forms of development assistance. While most GRPPs undertake periodic evaluations, these are usually formative evaluations for improving a specific aspect of the program's performance, or summative evaluations of program outcomes, rather than rigorous impact evaluations. However, impact evaluations can effectively complement or contribute to these formative and summative evaluations in providing accountability and in confirming that development funds have been spent wisely on effective interventions. While a rigorous impact evaluation of a GRPP at the program level would be extremely difficult, because of the diversity of its components and the resultantly complex causality and aggregation problems, impact evaluations of selected activities are feasible and are encouraged.

ADVANCE PLANNING FOR CONDUCTING IMPACT EVALUATIONS

15.3 As emphasized in Chapter 2 (paragraph 2.23), impact evaluation should be planned in advance, for several reasons. First, like all evaluations, impact evaluation needs to be based on accurate data,

which come from a mature and tested monitoring system. Second, to be the most credible, it is necessary to compare the welfare outcomes arising from the program with a counterfactual – what would have occurred in the absence of the program. A good technique for establishing the counterfactual consists of identifying one group receiving the intervention(s) and a similar control group not receiving the intervention(s), and then initiating early baseline data collection relating to *both groups* before the intervention(s) begin. This ensures that adequate information is available for the subsequent comparison of the situation with and without the intervention(s), once these have been in place long enough to have had an impact. This “*double-difference*” approach requires the early establishment of a research design and the early collection of baseline data, ideally even before the potential beneficiaries learn of the intervention and develop expectations that may affect their behavior.

15.4 If baseline data have not been collected, an impact evaluation can still be conducted by comparing the welfare outcomes of the group receiving the interventions with those of a control group, while attempting to control for other influences through statistical methods. This “*single difference*” approach also requires careful design and planning, and is also dependent on an established and tested monitoring system.

Standards and Guidelines

PLANNING FOR A PARTICULAR IMPACT EVALUATION

15.5 Impact evaluation needs to be planned carefully and employed selectively as one of several types of useful evaluations that can serve different purposes at different stages of a program. Impact evaluation would normally be considered more feasible and relevant after the program has reached a steady state in terms of financing, scope, and coverage. Because impacts are not usually manifest until after the passage of some time, the scope of an impact evaluation may cover only a subset of activities which have reached a certain stage of gestation or which were completed during a previous period of the program’s life.⁸⁶ If at any stage, a future quantitative impact evaluation of any intervention(s) is considered likely, the program should make provision for the collection of baseline data to provide the basis for comparison with the counterfactual.

86. The Consultative Group on International Agricultural Research (CGIAR) is the major GRPP that has conducted impact evaluations on its productivity-enhancing agricultural research. Some of the health programs, such as the Special Programme on Research and Training in Tropical Diseases (TDR), have also evaluated the impacts of their research on diseases of the poor such as onchocerciasis, leprosy, and malaria.

Draws on materials presented at the International Workshop on Impact Evaluation for Development, November 15, 2006

15.6 Impact evaluations of a specific intervention can be conducted as soon as it is judged likely that welfare outcomes have been realized. However, since impact evaluations are generally more costly than other forms of evaluation, the governing body should carefully consider the costs and benefits of conducting an impact evaluation of a given set of interventions at a given time. As a general guideline, for accountability and for assurance of continued relevance, the governing body should consider financing a comprehensive impact evaluation after 10 years of a program's life.⁸⁷

CONDUCT OF IMPACT EVALUATIONS ⁸⁸

15.7 Impact evaluations will usually be conducted for subsets of activities where impact is judged to be more measurable than for the program as a whole, or where there is a pressing need for an assessment of impact to influence design adjustments or decisions on replicability and scaling up. Impact evaluations will normally be conducted parallel to and not as part of a program-level evaluation. If the results of the impact evaluation are to be used in subsequent program-level evaluations, it is important that the sample for the impact evaluation be chosen in order to be representative. For example, the sample might include people from one village where the conditions seem favorable for high impact and from another where conditions are less favorable. Selecting an appropriate comparison group and avoiding selection bias are two of the major challenges in impact evaluation.⁸⁹

15.8 Good impact evaluations use a combination of quantitative impact evaluation, participatory impact evaluation, and theory-based (program logic) approaches. Qualitative participatory analysis helps to add context to and provide confirmation of findings derived from the other approaches. A theory-based approach helps to track the influences at different points in the results chain and to enhance understanding of when or why the program works well or not. Quantitative methods give an authoritative and credible indication of the relative impact of the program, compared with the counterfactual situation.

87. Several GRPPs that have existed for more than 15 years have never had an impact evaluation.

88. Impact evaluation is the subject of an ongoing working group of the OECD/DAC, and more detailed guidelines are expected.

89. *Selection bias* is the distortion that arises in a statistical analysis due to the methodology that was used to collect the samples. For instance, the beneficiaries of a certain intervention may be selected (or self-selected) on the basis of certain characteristics. If these are observed, then it is important to select a comparison group with the same characteristics. If these are unobserved, then only a randomized approach can in principle eliminate the selection bias.