ANNEX 3: DATA COLLECTION METHODS AND EVALUATION DESIGN

Notes on Data Collection Methods: While this module will not go into the methods of collecting data, it might be worth pointing out to the participants that there are different tools for collecting data. “Data collection and analysis methods for impact evaluation vary along a continuum. At the one end of this continuum are methods relying on random sampling; structured data collection instruments that fit diverse experiences into predetermined response categories; and statistical data analysis. These methods, generally associated with quantitative research, produce results that are easy to summarize, compare and generalize.”

At the other end of the continuum are methods typically associated with qualitative research. These methods are characterized by the following attributes:

- they tend to be open-ended and have less structured protocols (i.e., researchers may change the data collection strategy by adding, refining, or dropping techniques or informants)
- they rely more heavily on iterative interviews; respondents may be interviewed several times to follow up on a particular issue, clarify concepts or check the reliability of data
- they use triangulation to increase the credibility of their findings (i.e., researchers rely on multiple data collection methods to check the authenticity of their results)
- generally their findings are not able to generalize to any specific population; rather, each case study produces a single piece of evidence that can be used to seek general patterns among different studies of the same issue.

In between the two extremes, there is a number of possible evaluation methodologies combining different aspects (sample design, research protocol, data collection and data analysis) of the quantitative and qualitative approaches.

Evaluations can also rely on participatory methods. These tend to be closer to the qualitative than to the quantitative research approach. However, not all qualitative methods are participatory, and inversely, many participatory techniques can be quantified. The participatory approach is very much action-oriented. Thus, stakeholders themselves are responsible for collecting and analyzing the information, and for generating recommendations for change. The role of an outside evaluator is to facilitate and support this learning process. By combining these different approaches, one can enrich the design, as well as interpretation or explanation of outcomes measured by the evaluation.

1 This section is excerpted from the publicly available Web sites of the World Bank and North Central Regional Educational Laboratory. Please see Web sites for further details.


The qualitative methods most commonly used in evaluation can be classified in three broad categories:

- in-depth interviews
- observational methods
- document review

Three participatory methodologies and their associated tools and techniques, which are commonly used in participatory M&E are:

- Beneficiary Assessment (BA)
- Participatory Rural Appraisal (PRA)
- SARAR: Self-esteem, Associative strength, Resourcefulness, Action planning and Responsibility

There is a growing acceptance for the need to integrate the different approaches to evaluation. Impact evaluations, using survey data from statistically representative samples, may be better suited to assessing causality by using econometric methods or reaching conclusions that can be generalized. However, qualitative and participatory methods allow the in-depth study of selected issues, cases or events. The methods also provide critical insights into beneficiaries’ perspectives, the dynamics of a particular reform, or the reasons behind certain results observed in a quantitative analysis.

Integrating quantitative, qualitative and participatory methods can often be the best vehicle for meeting the program’s information needs. For example, qualitative methods can be used to inform the evaluation questions and the questionnaire design. The qualitative methods can be used to analyze the social, economic and political context in which a program or policy takes place. Similarly, quantitative methods can be used to inform qualitative data collection strategies, including sample design. Quantitative methods also apply statistical analysis to control for household characteristics and the socio-economic conditions of different study areas. These methods eliminate alternative explanations of the observed outcomes.

**Notes on Evaluation Design:**

Although this module does not explore evaluation design in detail, it is important to point out the need for an evaluation design.

**Quantitative Methods.** Essentially there are three possible types of evaluation design:
Experimental or randomized evaluation design: “This design involves gathering a set of individuals (or other unit of analysis) equally eligible and willing to participate in the program and randomly dividing them into two groups: those who receive the intervention (treatment group) and those from whom the intervention is withheld (control group). Experimental or randomized designs are generally considered the most robust of the evaluation methodologies. By randomly allocating the intervention among eligible beneficiaries, the assignment process itself creates comparable treatment and control groups that are statistically equivalent to one another, given appropriate sample sizes. This is a very powerful outcome because, in theory, the control groups generated through random assignment serve as a perfect counterfactual, free from the troublesome selection bias issues that exist in all evaluations. The main benefit of this technique is the simplicity in interpreting results—the program impact on the outcome being evaluated can be measured by the difference between the means of the samples of the treatment group and the control group.”

Quasi-experimental design: This design consists of constructing a comparison group using matching or reflexive comparisons. Matching involves identifying non–program participants comparable in essential characteristics to participants. Both groups should be matched on the basis of either a few observed characteristics, or a number of them that are known or believed to influence program outcomes. Matched comparison groups can be selected before project implementation (prospective studies) or afterwards (retrospective studies). The main advantage of evaluations using matching methods is that they can draw on existing data sources and are thus often quicker and cheaper to implement. The principal disadvantages are that the reliability of the results is often reduced as the methodology may not completely solve the problem of selection bias; and the matching methods can be statistically complex, thus requiring considerable expertise in the design of the evaluation and in analysis and interpretation of the results.

Non-experimental design: Non-experimental designs are types of studies characterized by the absence of a control or comparison group and are generally used only when one is trying to collect descriptive data. There are two commonly used non-experimental designs in evaluation research: (1) the post-test-only design and (2) the pre-test-posttest design. There are several key points to note about both of these non-experimental designs. First, while both can be used for descriptive purposes, neither can be used to claim that the intervention is better than any other intervention. The Pre-test-Post-test Design does allow one to judge the amount of gain made by the treatment group, but you cannot attribute this change to your intervention. It could be that time or other events that occurred during the intervening time period caused the gains between the first and second tests. Because of these problems, non-experimental designs are the designs of last choice. As with quasi-experimental methods, this evaluation design is relatively cheap and easy to implement since it can draw on existing data sources. However, it poses a number of difficulties. First, the reliability of results is often reduced as the methodology is less robust statistically. Second, the methodology has some statistical complexities that may require some expertise in the design of the evaluation and in the analysis and interpretation of results. Third, although it is possible to partially correct for selection bias, full correction remains as a challenge.
SAMPLE EVALUATION MATRIX:

<table>
<thead>
<tr>
<th>Goal:</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Objectives</td>
<td>Indicators</td>
<td>Measures</td>
<td>Sources/Data Collection Methods</td>
</tr>
<tr>
<td>Evaluation Question 1:</td>
<td>(1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evaluation Question 2:</td>
<td>(1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>