The work by Campbell and others on validity and threats to validity within experiments and other types of evaluations have left deep marks on the way researchers and evaluators have addressed methodological challenges in impact evaluation (see Campbell, 1957; Campbell and Stanley, 1963; Cook and Campbell, 1979; Shadish et al., 2002).

5.1. Different methodologies have comparative advantages in addressing particular concerns and needs

Validity can be broadly defined as the “truth of, or correctness of, or degree of support for an inference” (Shadish et al., 2002: 513). Campbell distinguished among four types of validity, which can be explained in a concise manner by looking at the questions underlying the four types:

- **Internal validity**: How do we establish that there is a causal relationship between intervention outputs and processes of change leading to outcomes and impacts?
- **Construct validity**: How do we make sure that the variables we are measuring adequately represent the underlying realities of development interventions linked to processes of change?
- **External validity**: How do we (and to what extent can we) generalize about findings to other settings (interventions, regions, target groups, etc.)?
- **Statistical conclusion validity**: How do we make sure that our conclusion about the existence of a relationship between intervention and impact variable is in fact true? How can we be sure about the magnitude of change?

Applying the logic of comparative advantages makes it possible for evaluators to compare methods on the basis of their relative merits in addressing particular aspects of validity. This provides a useful basis for methodological design choice; given the evaluation’s priorities, methods that better address particular aspects of validity are selected in favor of others. In addition, the logic of comparative advantages can support decisions on combining methods to be able to simultaneously address multiple aspects of validity.

We will illustrate this logic using the example of RCTs. Internal validity usually receives (and justifiably so) a lot of attention in impact evaluation, as it lies at the heart of the attribution problem; is there a causal link between intervention outputs and outcomes and impacts? Arguably, RCTs (see § 4.2.) are viewed by many as the best method for addressing the attribution problem from...
the point of view of internal validity. Random allocation of project benefits reduces the likelihood that there are systematic (observable and unobservable) differences between those that receive benefits and those that do not. However, this does not make it necessarily the best method overall. For example, RCTs control for differences between groups within the particular setting that is covered by the study. Other settings have other characteristics that are not controlled, hence there are limitations of external validity here.

To resolve this issue, Duflo and Kremer (2005) propose to undertake series of RCTs on the same type of instrument in different settings. However, as argued by Ravallion, “The feasibility of doing a sufficient number of trials—sufficient to span the relevant domain of variation found in reality for a given program, as well as across the range of policy options—is far from clear. The scale of the randomized trials needed to test even one large national program could well be prohibitive” (Ravallion, 2008: 19).

Another limitation of RCTs (also valid for other approaches discussed in § 4.2) lies in the realm of construct validity. Does the limited set of indicators adequately represent the impact of a policy on a complex phenomenon such as poverty? In-depth qualitative methods can more adequately capture the complexity and diversity of aspects that define (and determine) poverty than the singular or limited set of impact indicators taken into account in RCTs. Consequently, the latter have a comparative advantage in addressing construct validity concerns. However, a downside of most qualitative approaches is that the focus is local and findings are very context specific, with limited external validity. External validity can be adequately addressed by, for example, quantitative quasi- and non-experimental approaches that are based on large samples covering substantial diversity in context and people.

Theory-based evaluation provides the basis for combining different methodological approaches that have comparative advantages in addressing validity concerns. In addition, the intervention theory as a structure for making explicit causal assumptions, generalizing findings, and making in-depth analysis of specific assumptions can help strengthen internal, external, and construct validity claims.

To conclude:

- There is no single best method in impact evaluation that can always address the different aspects of validity better than others.
- Methods have particular advantages in dealing with particular validity concerns; this provides a strong rationale for combining methods.

5.2. Advantages of combining different methods and sources of evidence

In principle, each impact evaluation is in some way supported by different methods and sources of evidence. For example, even the quite technical quantitative approaches described in § 4.2 include other modes of inquiry, such as the research review to identify key variables that should be controlled for in, for example, a quasi-experimental setting. Nevertheless, there is a growing literature on the explicit use of multiple methods to strengthen the quality of the analysis.2 At the same time the discordance between the practice and “theory” of mixed-methods research (Bryman, 2006) suggests that mixed-methods research is often more an art than a science.

Triangulation is a key concept that embodies much of the rationale behind doing mixed-methods research and represents a set of principles to fortify the design, analysis, and interpretation of findings in impact evaluation.3 Triangulation is about looking at things from multiple points of view, a method “to overcome the problems that stem from studies relying upon a single theory, a single method, a single set of data … and from a single investigator” (Mikkelsen, 2005: 96). As can be deduced from the definition, there are different types of triangulation. Broadly, these are the following (Mikkelsen, 2005):
USE A MIXED-METHODS APPROACH: THE LOGIC OF THE COMPARATIVE ADVANTAGES OF METHODS

- Data triangulation—To study a problem using different types of data, different points in time, or different units of analysis
- Investigator triangulation—Multiple researchers looking at the same problem
- Discipline triangulation—Researchers trained in different disciplines looking at the same problem
- Theory triangulation—Using multiple competing theories to explain and analyze a problem
- Methodological triangulation—Using different methods, or the same method over time, to study a problem.

As can be observed from this list, particular methodologies already embody aspects of triangulation. Quantitative double-difference impact evaluation (see § 4.2.), for example, embodies aspects of methodological and data triangulation. Participatory impact evaluation approaches are often used to seek out and reconstruct multiple (sometimes contrasting) perspectives on processes of change and impact using diverse methods, often relying on teams of researchers with different disciplinary backgrounds (that may include members of target groups). Theory-based evaluation often involves theory triangulation (see chapter 3; see also Carvalho and White [2004], who refer to competing theories in their study on social funds). Moreover, it also allows for methodological and data triangulation by relying on different methods and sources of evidence to test particular causal assumptions.

Discipline triangulation and theory triangulation both point to the need for more diversity in perspectives for understanding processes of change in impact evaluation. Strong pleas have recently been made for development evaluators to recognize and make full use of the wide spectrum of frameworks and methodologies that have emerged from different disciplines and that provide evaluation with a rich arsenal of possibilities (Kanbur, 2003; White, 2002; Bamberger and White, 2007). For example, when doing impact evaluations, evaluators can benefit from approaches developed in different disciplines and subdisciplines. Neo-institutionalist economists have shown ways to study the impact of institutions as “rules of the game” (see North, 1990), and interventions such as policies can be considered as attempts to establish specific rules with the expectation (through a “theory of change”) of generating certain impacts (Picciotto and Wiesner, 1997). In addition, the literature on behavioral and social mechanisms (see appendix 10; see also chapter 6) provides a wealth of explanatory insights that help evaluators better understand and frame processes of change triggered by interventions.

A good methodological practice in impact evaluation is to encourage applying these principles of triangulation as much as possible.

Advantages of mixed-methods approaches to impact evaluation are the following:

- A mix of methods can be used to assess important outcomes or impacts of the intervention being studied. If the results from different methods converge, then inferences about the nature and magnitude of these impacts will be stronger. For example, triangulation of standardized indicators of children’s educational attainments with results from an analysis of samples of children’s academic work yields stronger confidence in the educational impacts observed than either method alone (especially if the methods employed have offsetting biases).

- A mix of methods can be used to assess different facets of complex outcomes or impacts, yielding a broader, richer portrait than one method alone can. For example, standardized indicators of health status could be mixed with onsite observations of practices related to nutrition, water quality, environmental risks, or other contributors to health, jointly yielding a richer understanding of the intervention’s impacts on targeted health behaviors. In a more general sense, quantitative impact evaluation techniques work well for a limited set of pre-established variables (preferably determined and measured ex ante) but less well for capturing unintended, less expected (indirect) effects of interventions. Qualitative methods or descriptive (secondary) data
analysis can be helpful in better understanding the latter.

• One set of methods could be used to assess outcomes or impacts and another set to assess the quality and character of program implementation, including program integrity and the experiences during the implementation phase.

• Multiple methods can help ensure that the sampling frame and the sample selection strategies cover the whole of the target intervention and comparison populations. Many sampling frames leave out important sectors of the population (usually the most vulnerable groups or people who have recently moved into the community), while respondent selection procedures often under-represent women, youth, the elderly, or ethnic minorities. This is critical because important positive or negative impacts on vulnerable groups (or other important sectors) are completely ignored if they do not even get included in the sample. This is particularly important (and frequently ignored) where the evaluation uses secondary data sets, as the evaluator often does not have access to information on how the sample was selected.

• Multiple methods are needed to address the complementary questions of average effect and distribution of costs and benefits of an intervention (see § 5.3.)

Appendix 11 presents four interesting examples of impact evaluations that are based on a mixed method perspective:

• Case 1: Combining qualitative and quantitative descriptive methods—Ex post impact study of the Noakhali Rural Development Project in Bangladesh
• Case 2: Combining qualitative and quantitative descriptive methods—Mixed-methods impact evaluation of International Fund for Agricultural Development projects in Gambia, Ghana, and Morocco
• Case 3: Combining qualitative and quantitative descriptive methods—Impact evaluation: agricultural development projects in Guinea
• Case 4: A theory-based approach with qualitative methods (GEF, 2007).

5.3. Average effect versus distribution of costs and benefits

Sometimes policy makers and stakeholders are concerned with the question of whether an intervention (for a specific context and group of people) has been effective overall. This is typically a question that can be addressed by using (quasi) experimental evaluation techniques. However, another important question, one that might not be easily answered with these techniques, is whether and how people are differently affected by an intervention. This question can be answered by using regression analysis. A regression model can incorporate different moderator variables (e.g., through modeling interaction effects) to analyze to what extent important characteristics co-determine outcome variables. In addition, many qualitative methods such as those used for case studies can help evaluators study in detail how interventions work differently in different situations. From a methodological design perspective, a mixed-methods study combining quasi-experimental survey data with a limited number of in-depth, semistructured interviews among different types of people from the target population is an example of a potentially good framework to provide credible answers to both questions (see box 5.1.).

When talking about the issue of distribution of costs and benefits of an intervention, it is useful to distinguish between different levels or foci. First, one should consider the issue of outreach or coverage. Who are the people (individuals, households, and communities) directly affected by an intervention? Sometimes this question can be answered in a relatively straightforward manner, such as when the intervention is clearly delineated and targeted to a specific group of people (e.g., a training program). In other cases (e.g., a tax cut or construction of a road), coverage or outreach, or indeed the delineation of the group of people affected by the intervention, is not that easy to determine. In the last case, the issue of delineation is closely linked to the second level, how an intervention has different effects on
Important to note is that an analysis of the distribution of costs and benefits as a result of an intervention—distinguishing among coverage, effects on those who are directly affected, and indirect effects—cannot be addressed with one particular method. If one is interested in all these questions, then inevitably one needs a framework of multiple methods and sources of evidence. For example, descriptive analysis of survey data can help to map coverage, quasi-experiments can help to assess attribution of change among those directly affected, and case studies and survey data analysis can help to map indirect effects over time.

Consider the example of an intervention that provides monetary incentives and training to farmers to promote land use changes leading to improved livelihoods conditions. We could use the following methods in the impact evaluation:

- A randomized experiment could be used to assess the effectiveness of different incentives on land use change and/or socio-economic effects of these changes (potentially strengthens internal validity and increases external validity of findings).
- Survey data and case studies could tell how incentives have different effects on particular types of farm households (potentially strengthens internal validity and increases external validity of findings).
- Semistructured interviews and focus group conversations could tell us more about the nature of effects in terms of production, consumption, poverty, etc. (potentially enhances construct validity of findings).

Often, impact evaluation is about level two—determining the effects on those that are directly targeted by/participating in the intervention. In those cases, it is often assumed that level one (targeting, outreach) is fully known and mapped. In other cases, level one—outreach and coverage or indeed the determination of the scope of direct effects of an intervention on the population at risk—is the great “unknown” and should be a first priority in an impact evaluation exercise. Level three—indirect processes of change induced by an intervention, with potentially important implications for the distribution of costs and benefits among target populations and beyond—is often outside the scope of impact evaluations (see Ravallion, 2008).

Key message

Use a mixed-methods design. Bear in mind the logic of the comparative advantages of designs and methods. A mix of methods can be used to assess different facets of complex outcomes or impacts, yielding more breadth, depth, and width in the portrait than one method alone can. One set of methods could be used to assess outcomes or impacts and another set to assess the quality and nature of intervention implementation, thus enhancing impact evaluation with information about program integrity and program experiences. It is important to note that an analysis of the distribution of costs and benefits of an intervention—distinguishing among coverage, effects on those directly affected, and indirect effects—cannot be addressed with one particular method. Answering these questions requires a framework of multiple methods and sources of evidence.