Smart AL

Co-funded by the Erasmus+ Programme of the European Union

Executive Master

"Territorial Planning and Economic Development of Rural Areas"

Booklet of Didactic Material

Module 8

Instruments and policy evaluation methods

September 2021

CIHEAM MONTPELLIER · MUSE MONTPELLIER UNIVERSITÉ D'EXCELLENCE · Universiteti Bujqësor i Tiranës · UK FAN S. NOLI · Universiteti Europian i Tiranës · UNIVERSITY OF HOHENHEIM · Université Paul Valéry Montpellier 3

Module developed by:

- Dr. Dorjan Marku, University Fan S. Noli, Korçë, Albania (UNKO)
- Prof. Dr. Manfred Zeller, University of Hohenheim, Germany (UHOH)
- Prof. Dr. Reiner Doluschitz, University of Hohenheim, Germany (UHOH)

Project implemented under the coordination of Magali MAIRE, International Center for Advanced Mediterranean Agronomic Studies - Mediterranean Agronomic Institute of Montpellier, France (CIHEAM-IAMM), with the support of Virginie AVIGNON, CIHEAM-IAMM.

Table of contents

List of figures:

List of tables:

Preface

This booklet is prepared by PhD. Dorjan Marku, lecturer at "Fan S. Noli" University of Korca, Agribusiness Department, in close collaboration and consultation with Professor Dr. Reiner Doluschitz, evaluation expert and lecturer at the University of Hohenheim, and Professor Dr. Manfred Zeller, expert and lecturer of the module "Monitoring and Evaluation of Development Projects and Policies" at the University of Hohenheim.

The module "Instruments and Policy Evaluation Methods" is prepared in the framework of the Smart AL Project (Erasmus Plus), as part of the teaching modules for the implementation of the Master Program "European Innovations for a Sustainable Management of Albanian Territories, rural areas and agriculture: Instruments, policies, strategies".

The content of this booklet can be used only by the partner institutions involved in the project, and in the meantime, it may serve as a reference for the teachers involved and the students that will be enrolled in the Master Program.

The objectives of this booklet are to provide the reader with:

- Disputed notions of rural, regional, territorial and development terms
- A basic understanding of the new challenges of territorial development and the tools to be applied in order withstand to these changes
- Different result-oriented instruments that are applied in order to facilitate the management of trade-offs by policy makers in terms of regional and rural development
- Knowledge on the main principles of monitoring and evaluation, indicators and practical guidelines for designing and implementing impact evaluations
- Knowledge on the features of successful monitoring and evaluation design and the log frame approach applied on the preparation and implementation of projects, programs and policies.
- Knowledge about impact evaluation methodologies that can be used to estimate the impact of programs and policies, as a form to increase efficiency

GENERAL INTRODUCTION

The module "Instruments and Policy Evaluation Methods" focuses on the public policy instruments, the identification of the policy answer to different issues that are related to rural and peri-urban areas, and most important on the impact evaluation and various methodologies that may be used in order to estimate the impact of policies, programs and projects applied on rural and urban territories and communities.

Nowadays, regions and their communities are facing important challenges such as globalization, climate change, etc. In this context, policy makers need to design more efficient policies in terms of regional development, and manage better the policy tradeoffs, such as: ensuring that the instruments applied for the regional development to be sufficiently flexible and in the meantime to ensure policy stability and accountability; To find out a balance between performance, compliance and administrative costs. On the other side, it is important to underline that different instruments that are used for the development and the sustainable management of the territories should reflect territorial specificities and be adapted depending on the different problem context. This flexibility can respond more effectively to the different needs that different regions have and ensure the efficiency of the resources that are used.

Capacity gaps may have a direct impact on the development of regions and inequalities that exist between them. So, another very important dimension that the module emphasizes is related to the capacity development during the implementation of projects, programs and policies. In order to sustain changes that derive at the end of a project or policy, organizations and governments should put more efforts in enhancing their systems, structures and institutions. Hence, if we refer to the institutional environment and the issues related to territorial and regional development in Albania, despite the reinforcement of skills and abilities, capacity building should also target institutional and financial capacities. Consequently, capacity building needs to be understood as a "learning by doing" process.

In addition, setting monitoring and evaluation systems in the early stages of the project, program or policy design process is necessary for increasing efficiency. Monitoring and evaluation can help policymakers and organizations extract useful information from past experiences and ongoing activities, helping them to adapt the objectives of the projects, programs and policies to better fit the needs of the rural regions and cities, having a positive impact on their communities. Also, impact evaluations allow officials and the human resources involved on the projects, to allocate the resources that are necessary for defining the evaluation methodologies and produce relevant data.

PART I. OVERARCHING ISSUES OF RURAL DEVELOPMENT ASSESSMENT

CHAPTER I. ISSUES OF RURAL DEVELOPMENT AND TERRITORIES : CONSENSUS, CONTROVERSIES AND UNCERTAINTIES

When it comes to rural and peri-urban areas, the development of these territories and policies, there are some very important issues that we should pay attention.

1. According to World Bank statistics, rural areas are constantly changing, and in this context these territories represent the major part of world's surface area (37.7 % of the total agricultural land, 3.4 billion inhabitants, 46 % of the global population) making them extremely important for present and future events world widely (World Bank, 2001);
2. In terms of development, their diversity within regions and countries and the increased influence from urban territories, has oriented them toward a strong competition;
3. They contain most of the resources necessary for human existence, and because of the scarcity of these resources, they are on focus of public policies.

Based on the issues mentioned above, the development of these territories are central at the agenda to policy makers, researchers and interest groups, because of their necessary for thorough analysis and future studies. In addition, on the sections below will be reviewed some very important approaches related to territorial and regional development in rural and peri-urban areas, their policies and respective scopes (Torre A., Wallet F., 2016).

Nowadays rural areas are subject to different changes and especially are facing two fundamental types of change:

1. ***Rural areas are constantly influenced from cities and urban populations***
2. ***The abundancy of natural resources located in rural areas is a determinant for future development strategy and development policies***

1.1 NOTIONS AND DEFINITIONS IN RURAL AND REGIONAL STUDIES

Usually, the notions over the terms, rural, regional, territorial and development are subjects of debate and controversy in the contemporary literature and in the context of the criteria's that are used by different organizations and governments in order to define them.

Specifying the distinction between “rural” and “urban” or giving a clear-cut definition of the term “rural” it is not an easy task (Mormont 1990), but the uncertainty that characterizes current transformations may give rise to a *productive debate.*

The frontier between rural and urban domains, often mentioned in reference to the city-country relationship, has weakened or even disappeared as a result of a twofold process. On the one hand, areas traditionally devoted to agriculture have been urbanized, with a dramatic increase in the number of buildings and individual houses encroaching upon open spaces. On the other hand, rurality and agriculture are making their way into cities, as demonstrated by the growing success of locally based farming, local food systems and transition towns (Reid et al. 2012), and even urban agriculture (Despommier 2010).

The notions of **“regional development”** and **“territorial development”** are considered to be very close to the term “local development”, generally applied to small, infra-regional portions of territory. Even though the generalized use of these expressions raises questions: are they identical, opposed, or substitutable? And, above all, this raises the question of how the concepts of region and territory are defined.

The term *“regional”* refers to two relatively distinct definitions. The first, mostly used in an administrative sense by the regional authorities, refers to administrative regions (e.g. the Centre region in France, or the Tuscany region in Italy). The second, pertains to the “geographical” dimensions of development or growth (e.g. Isard 1956). It encompasses questions relating to the “local”, the region, the location of activities or people, as well as the wealth and competitiveness of certain portions of space or nations.

It refers to the concept of *territory*, whose emergence was slow and sometimes controversial in this field of analysis. Sack (1986), gives a conventional definition referring to the concept: “Territoriality will be defined as the attempt by an individual or a group to affect, influence or control people, phenomena, and relationships, by delimiting an asserting control over a geographic area. This area will be called the territory” (Torre A., Wallet F., 2016). As a result, territories are permanent constructs with moving boundaries, and are constituted through oppositions between and compromises among local actors.

Based on these definitions, the concept of regional development refers to the processes that occur within the institutional borders of the region, whereas that of territorial development pertains to a construction of territorialities by local populations (Mollard et al. 2007), in relation with policy directives or more general incentives.

1.2 APPROACHES IN TERMS OF REGIONAL BALANCE AND GROWTH:

1.2.1 Homothetic Growth and Economic Base Theory

*The **homothetic growth and economic base analysis*** approaches seeks to balance the interests and gains from the development process enjoyed by different local actors and to draw up principles that will enable the various stakeholders to obtain maximum satisfaction.

- ✓ Included in the first group (***homothetic growth***, *based on stakeholder preferences*) are the approaches that underline the neoclassical theory, approaches which envisage a form of ***homothetic growth*** based on capital and labor inputs, subsequently extended to a third input of a more technological nature (Solow 2000).
- ✓ ***Economic base analysis*** (Alexander 1954; Sombart 1916) also advocates balanced development. It rests on the idea that regional economies can be divided into two main components:
  - i. A "basic sector", which produces goods and services for export and fosters regional development by capturing revenue from external trade;
  - ii. A domestic sector, whose production is for local consumption.

Development then requires expansion of the basic sector that gives rise to a Keynesian multiplier effect on the local economy as a whole. The increasing income of those who work in this sector then generates a rise in their consumption levels and, as a result, a development of the domestic production sector.

1.2.2 The unbalanced approaches

*The **unbalanced approaches*** consist of approaches whereby the compromises reached among local actors are purely temporary and development processes generate interregional inequalities that are difficult to reduce. These approaches consider that development plays an important role in increasing disparities between regions or territories.

1.2.3 The systemic approach

*The **systemic approach*** is linked to the systemic nature of the relationships between actors who, together, belong to one territory and shape it through their cooperation and common projects (Zeller and Nielson, 2013).

The first is ***Porter's analysis*** (Porter 1985 and 1990), because of its broad impact. Porter considers that the competitive advantage of a region or a territory depends on four main factors that must be exploited in order for the region in question to gain a lead over its competitors (Porter 1985):

- ✓ The strategies, structures and the rivalry between firms;

- ✓ The state of demand;
- ✓ The geographical relations between similar firms;
- ✓ The state of production resources or factors (traditional or skill-related)

The second concept, equally important, is that of ***industrial districts.*** Present in the works of Alfred Marshall as early as 1920, districts were rediscovered in the seventies by the Italian economists (Brusco 1982). Becattini (1990) defines an ***industrial district*** as "a socio-territorial entity which is characterized by the active presence of a community of people and a population of firms in one naturally and historically bounded area."

1.3 WHAT WILL BE THE FUTURE FOR RURAL AREAS? SCENARIOS FOR POSSIBLE DEVELOPMENT PATHS

Throughout the 20th century, rural areas has undergone profound changes, raising different questions for the future evolution of these areas. These profound changes have placed them in a relatively unstable and shifting position, which contrast the period when agriculture was undoubtedly the core activity of these regions and the primary source of their wealth, on which ultimately all other production and activities depended.

For the development of rural and peri-urban areas five important scenarios should be examined (Wiggins, 2015):

- ➢ The possible preservation of agricultural activities and their prevalence;
- ➢ The unrelenting rise of urbanization and peri-urbanization;
- ➢ The role of industry and business in rural areas;
- ➢ The development of services to individuals and of the residential economy;
- ➢ The coexistence of different land uses, and competition between them, in certain areas;

CHAPTER II. CAPACITY BUILDING FOR MONITORING AND EVALUATION IN PUBLIC ADMINISTRATION

2.1 KEY CONCEPTS IN CAPACITY BUILDING

Good M&E is dependent on good planning. If the monitoring and evaluation of capacity building is to be effective it is important to know what the purpose of capacity building is, who are the providers and recipients of capacity building, and whose perspectives we are interested in. Only then can the various M&E alternatives be considered.

One of the key challenges for anyone involved in the M&E of capacity building is to agree what is meant by the term. This is not easy, as there are many different definitions, some of which are contradictory.

At its most basic **capacity** can be understood as 'the ability of people, organizations and society as a whole to manage their affairs successfully' (OECD 2006, p8). **Organizational capacity** can be defined as 'the capability of an organization to achieve effectively what it sets out to do. (Fowler et al 1995, p4).

The capacity of an individual, an organization or a society is not static. It changes over time, and is subject to both internal and external influences. Many of these changes are unplanned. For example, an organization can lose capacity if key individuals leave or change positions within that organization. However, **capacity development** can be seen as a more deliberate process whereby people, organizations or society as a whole create, strengthen and maintain capacity over time.

2.2 CAPACITY BUILDING FOR WHAT?

At the organizational level, capacity building is carried out for a variety of different purposes. Broadly, these can be divided into two major groups (Zeller and Nielson, 2013).

i. **Technical capacity building** is aimed at addressing a specific issue concerning an organization's activities. Technical capacity building would not normally be expected to involve an organization in a fundamental process of change, and would be unlikely to touch on the culture, vision, values or other core elements of that organization.

ii. **General capacity building** is provided to help organizations develop their own capacity to better fulfil their core functions, and achieve their own mission. This type of capacity development can be slow, complex and continuous, and can require in-depth reflection on an organization's culture, values and vision. The ultimate goal of such work is to improve the organization's overall performance and its ability to adapt itself within a changing context.

A **theory of change** at the organizational level might cover the different ways in which organizations change. Reeler (2007) describes three different kinds of change and argues that the type of change considered has profound implications for M&E:

1. **Emergent change** describes the day-to-day changes that are brought about by individuals, organizations and societies adjusting to changing circumstances, trying to improve what they know and do, building on what is already there, and constantly learning and adapting.

2. **Transformative change** occurs when an organization becomes stuck or goes through a period of crisis, either through natural processes or external shocks. In this case the change process is one of unlearning inappropriate ideas and values and adopting new ones in order to create a new situation.
3. **Projectable change** is the kind of change that can be planned in advance, and made the focus of a specific project or piece of work. It is more about working to a plan to build on or negate visible challenges, needs or possibilities.

On the other hand, different theories of change can also be used to describe how organizational change contributes to wider aims and objectives. Ortiz and Taylor (2008) stress the importance of organizations having a clear understanding of how change happens. They argue that this means understanding the demands or needs of primary stakeholders, and the conditions required to support the emergence of change, as well as understanding the broader socio-economic environment.

2.3 RESOURCES FOR MONITORING AND EVALUATION

Inadequate resources lead to poor quality of monitoring and evaluation. To ensure effective and quality monitoring and evaluation, it is critical to set aside adequate financial and human resources at the planning stage. The required financial and human resources for monitoring and evaluation should be considered within the overall costs of delivering the agreed results and not as additional costs.

Financial resources for monitoring and evaluation should be estimated realistically at the time of planning for monitoring and evaluation.

Human resources are critical for effective monitoring and evaluation, even after securing adequate financial resources.

2.4 CAPACITY BUILDING AS A *LEARNING BY DOING* PROCESS

The use of different techniques and mechanisms to guide, co-ordinate, and align priorities for regional development require the existence of certain capacities at the different levels of government involved in the investment cycle.

Capacity building is also a “learning-by-doing” process in which supra-national, national and subnational stakeholders learn by repeated interactions.

2.5 MONITORING AND EVALUATION OF CAPACITY BUILDING

If organizations are to carry out effective M&E around capacity building, a key first question to address is what is the purpose of that M&E? The usual answer to this is a

combination of accountability and learning in order to improve performance. But it is not always that simple. This is for two main reasons:

- ✓ M&E carried out to learn and improve performance will not necessarily meet the needs of accountability, and vice versa.
- ✓ There are likely to be competing demands on M&E within and across different organizations.

The challenge is often to reconcile all these competing demands. In many cases this can best be done by ensuring that monitoring and evaluation meets the needs of the primary stakeholders – the providers and recipients of capacity building. It is important to note the difference between **M&E of capacity** and **M&E of capacity building**.

Depth of measurements (*deciding how far to measure*)

One key decision is how far to go with M&E. For example, is it enough for a capacity building provider to show that its efforts have helped an organization (or individual) improve capacity, or should providers go further and measure the wider effects of these changes? To some extent, this depends on the purpose of the capacity building support. But it also depends on what is meant by measuring change. For example, some organizations attempt to **measure** capacity through the use of organizational assessment tools.

2.6 HOW TO CARRY OUT A CAPACITY ASSESSMENT

A capacity assessment is a useful way to look inside the "country capacity system" and initiate a focused dialogue between resource partners and national and local actors about meaningful interventions to strengthen national capacities to drive development.

Capacity assessments usually begin with the following questions:

- ✓ Capacity for what?
- ✓ Capacity for whom?

The reason for carrying out assessments is not "to know everything about everything", but to conduct an appropriate level of analysis to support decisions regarding capacity development for the initiative under consideration.

2.7 SELECTED TOOLS FOR CAPACITY DEVELOPMENT PRACTITIONERS

FAO recommends that, as a start, a capacity assessment should include three basic tools as shown in figure 3.

1. **A Problem Tree to answer the question 'Capacity for what?**
   A stakeholder mapping to understand what actors might need to be influenced, empowered or included and to answer the question 'Capacity for whom?'

A Capacity Questionnaire explores the three dimensions of the country capacity system: the enabling environment, organizations and individuals.

2. **What is the best way to get there?** This will compare the future with the present situation, and identify how to get from the current capacity to the desired future capacity.

CHAPTER III. AGRICULTURAL POLICY INSTRUMENTS

Different rural development policies have existed for decades and their success has always been acknowledged by local actors, they have nevertheless been subject to many shifts in vision and strategy. Both the policies and the concept of rural development have evolved with economic circumstances, been discussed in the same debates and have undergone the same reversals. They have changed in parallel with the shifts in focus from large-scale farming production to recognition of the multifunction's of agriculture, or with transitions from centralized decision making to greater inclusion of the various users of rural areas and even greater consideration for social criteria and ecological and environmental variables.

The need for rural policy derives from observing the differential incidence and persistence of poverty between rural and urban areas in the least and most developed countries alike, the pervasiveness of environmental degradation associated in part with the very same determinants of poverty and negative spillovers from metropolitan areas. Determinants of these problems can be traced back to the structural features of rural areas (distance, dispersion, resource-based activities, incomplete property rights, inequality in the distribution of assets, etc.), the pervasiveness of market failures for a significant share of households (particularly for credit, insurance and information, as well as high transaction costs in accessing product and factor markets), serious gaps in agrarian institutions essential for productivity and welfare, lags in the inter-sectoral reallocation of resources, a lack of coordination to escape regional low level equilibrium traps, pro-urban policy biases, and a lack of bargaining power for the rural poor.

3.1 UNDERGOING CHANGES OF PUBLIC POLICIES IN TERMS OF RURAL DEVELOPMENT

For a long time centered on agricultural issues, rural policies have, since the 1990s, undergone important shifts. They are increasingly diversified and oriented towards ensuring better management, exploitation and preservation of local resources, providing support for business and commercial activities in the secondary and tertiary sectors,

maintaining or increasing populations in rural areas (residential logic) and improving the organization of territories. Up to date policies of rural development are being implemented almost everywhere in the world, taking various forms according to the type of rural areas and their preferences in terms of development: mass farming production, production of high-quality products, residential development or tourist activities, for example.

Four main categories of public intervention can today be identified, each an important instrument in the toolkit used by policymakers interested in the development of rural areas Reimers (2018):

i. Policies for the conservation and development of localized resources such as land (through the promotion of agriculture and forestry) and of natural or built heritage;
ii. Interventions promoting economic activities, whether through industrial enterprises or commercial activities in the secondary and tertiary sectors, including tourism;
iii. Policies to encourage rural populations to remain in rural areas or attract new migrants to these areas, via a more residential approach;
iv. Actions facilitating territorial land planning and management, including through the creation of local facilities and institutions for rural land management, or even through the reorganization of towns and municipalities (related to the decentralization process).

3.2 THREE IMPORTANT CONSTANT SOLUTIONS FOR SELECTING SPECIFIC DEVELOPMENT PATHS AND INDIVIDUAL INSTRUMENTS

All the different aspects related to requirements and needs for the development of rural territories and their communities, and the various factors influencing farms and enterprises along the agricultural value chain, lead us to a very important prognosis: It is not possible to come up with a one-size-fits all solution for any country, region or farm enterprise. Nonetheless, it is possible to highlight a number of constants that play an important role in selecting a specific development path and applying the individual instruments:

i. **Creating a favorable investment cli-mate for farming families and rural areas**
ii. Investing in **vocational training opportunities for young people that are accessible to all**.
iii. The economic prosperity of the agricultural sector and rural areas in general is contingent on a **higher respect for agriculture by society as a whole**.

3.3 AGRICULTURAL POLICY INSTRUMENTS

This section provides an overview of the agricultural policy instruments that fall into six subject areas: market and pricing policy, agricultural and infrastructure policy,

agricultural education policy, agri-environmental policy, agricultural social policy, and consumer protection and animal welfare.

PART II. PRINCIPLES OF POLICY EVALUATION

CHAPTER IV. GENERAL PRINCIPLES OF IMPACT EVALUATION

4.1 EVIDENCE BASED POLICY MAKING: REASONS FOR EVALUATION

Development programs and policies are typically designed to change outcomes such as raising incomes, improving learning, or reducing illness. More commonly, program managers and policy makers focus on measuring and reporting the inputs and immediate outputs of a program, how much money is spent, how many people participate in an employment program, rather than on assessing whether programs have achieved their intended goals of improving outcomes.

Impact evaluations are part of a broader agenda of *evidence-based policy making.* This growing global trend is marked by a shift in focus from inputs to outcomes and results. Monitoring and evaluation are at the heart of evidence-based policy making. They provide a core set of tools that stakeholders can use to verify and improve the quality, efficiency, and effectiveness of policies and programs at various stages of implementation, in other words, to focus on results.

4.2 DEFINITIONS OF MONITORING AND EVALUATION

Impact evaluation is one of many approaches that support evidence-based policy, including monitoring and other types of evaluation. Before conducting an evaluation, it is crucial to ask (World Bank, 2016):

- ✓ Why are the evaluation done?
- ✓ How will the results be used?
- ✓ Who will be influenced by the findings?

The answer to these questions should determine:

- ➢ The process of evaluation
- ➢ Steps needed to be taken before quantitative data collection in the field is contemplated
- ➢ Designs of (quantitative) evaluations meet different needs of decision makers

Monitoring is a continuous process that tracks what is happening within a program and uses the data collected to inform program implementation and day-to-day management and decisions.

Evaluations are periodic, objective assessments of a planned, ongoing, or completed project, program, or policy. Evaluations are used selectively to answer specific questions related to design, implementation, and results.

4.3 DIFFERENCES BETWEEN MONITORING AND EVALUATION

Monitoring: the continuous assessment of project implementation in relation to schedules, input use, infrastructure, and services.

- ✓ Provides continuous feedback
- ✓ Identifies actual or potential successes/problems

Evaluation: the periodic assessment of a project's relevance, performance, efficiency, and impact (both expected and unexpected) in relation to stated objectives.

- ✓ Interim evaluations as a first review of progress, prognosis of likely effects, and as a way to identify necessary adjustments in project design.
- ✓ Terminal evaluations on project's effects and potential sustainability
- ✓ Evaluation gap

Table 1. Differences between monitoring and evaluation

Variables	Monitoring	Evaluation
Frequency	Periodic, regular	Episodic
Main action	Keeping track, oversight	Assessment
Basic purpose	Improve efficiency, adjust to work plan	Improve effectiveness, impact, future programming
Focus	Inputs, outputs, process outcomes, work plans	Effectiveness, relevance, impact, cost-effectiveness
Information sources	Routine systems, field observations, progress reports, rapid assessment	Same as for monitoring, as well as surveys and studies
Undertaken by	Program managers, community workers, community (beneficiaries), supervisors, funders	Program managers, supervisors, funders, external evaluators, community (beneficiaries)
Reporting to	Program managers, community workers, community (beneficiaries), supervisors, funders	Same as for monitoring, as well as policymakers

Source: UNICEF (1991:4)

4.4 DIFFERENT TYPES OF EVALUATIONS AND QUESTIONS THAT TRY TO ADDRESS

Different types of evaluations are appropriate for answering different kinds of questions. There is no "one size fits all" evaluation template to put against the variety of questions. It is important for managers to understand what they want to know from evaluations. Figure 1 depicts seven broad evaluation strategies that can be used to generate evaluation information. Each is appropriate to specific kinds of evaluation questions, and each will be briefly reviewed.

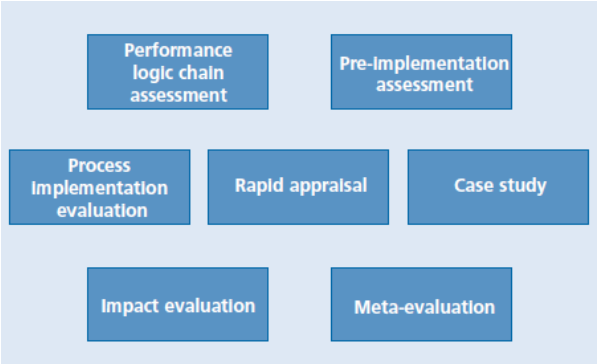

Figure 1. Seven Evaluation strategies

Source: The World Bank (2004): Ten steps to a results-based monitoring and evaluation system

4.5 IMPACT EVALUATION FOR POLICY DECISIONS AND MEASURES

Impact evaluations are needed to inform policy makers on a range of decisions, from curtailing inefficient programs, to scaling up interventions that work, to adjusting program benefits, to selecting among various program alternatives. They are most effective when applied selectively to answer important policy questions, and they are often applied to innovative pilot programs that are testing an unproven, but promising approach.

Impact evaluations can be used to explore different types of policy questions. The basic form of impact evaluation will test the effectiveness of a given program. In other words, it will answer the question, is a given program/project or intervention effective compared to the absence of them? This type of impact evaluation relies on comparing a treatment group that received the innovation, program, or policy to a comparison group that did not in order to estimate effectiveness. The core challenge in an impact evaluation is to construct a comparison group that is as similar as possible to the treatment group. The degree of comparability between treatment and comparison groups is central to the evaluation's *internal validity* and is therefore fundamental to assessing a program's causal impact.

Evaluations can also be used to test the effectiveness of program implementation alternatives. In this type of evaluation, two or more approaches or design features within a program can be compared with one another to generate evidence as to which is the most

cost-effective alternative for reaching a particular goal. These types of evaluations allow decision makers to choose among implementation alternatives, and can be very useful for enhancing program performance and saving costs.

4.6 DECIDING WHETHER TO CARRY OUT AN IMPACT EVALUATION

Not all programs warrant an impact evaluation. Impact evaluations should be used selectively when the question being posed calls for a strong examination of causality.

1. The first question to ask is, what is at stake? Will evidence about the success of the program, program modality, or design innovation inform important decisions? These decisions often involve budgetary allocations and program scale. If there are limited budget implications or if the results will affect only a few people, it may not be worth doing an impact evaluation.
2. Next question may be: *Is there any evidence that show if the program works?* Is there evidence available from similar programs under similar circumstances? If no evidence is available about the potential of the type of program being contemplated, then it may be needed to start out with a pilot that incorporates an impact evaluation.
3. A final question to ask is, *do we have the resources necessary for a good impact evaluation?* These resources concern technical elements such as appropriate data and time, financial resources to carry out the evaluation, as well as institutional resources with respect to the teams involved and their interest in and commitment to building and using causal evidence.

4.7 PROSPECTIVE VS RETROSPECTIVE IMPACT EVALUATION

Impact evaluations can be divided into two categories: prospective and retrospective.

Prospective evaluations are developed at the same time as the program is being designed and are built into program implementation. Baseline data are collected before the program is implemented for both the group receiving the intervention (known as the *treatment group*) and the group used for comparison that is not receiving the intervention (known as the *comparison group*).

Retrospective evaluations assess program impact after the program has been implemented, looking for treatment and comparison groups ex post. By contrast, in retrospective evaluations, the team that conducts the evaluation often has such limited information that it is difficult to analyze whether the program was successfully implemented and whether its participants really benefited from it. Retrospective evaluations using existing data are necessary to assess programs that were established in the past.

4.8 EFFICACY AND EFFECTIVENESS STUDIES

The main role of impact evaluation is to produce evidence on program performance for the use of government officials, program managers, civil society, and other stakeholders. Impact evaluation results are particularly useful when the conclusions can be applied to a broader population of interest.

In the early days of impact evaluations of development programs, a large share of evidence was based on ***efficacy studies:*** studies carried out in a specific setting under closely controlled conditions to ensure fidelity between the evaluation design and program implementation. By contrast, ***effectiveness studies*** provide evidence from interventions that take place under normal circumstances, using regular implementation channels, and aim to produce findings that can be generalized to a large population.

Key Concept: *Efficacy studies* assess whether a program *can* work under ideal conditions, while *effectiveness studies* assess whether a program *does* work under normal conditions.

4.9 COMPLEMENTARY OF DIFFERENT EVALUATION APPROACHES (*MONITORING PROGRAM IMPLEMENTATION, EX ANTE SIMULATIONS, MIXED METHODS, PROCESS EVALUATIONS*)

As noted, impact evaluations answer specific cause-and-effect questions. Other approaches, including close monitoring of the program, as well as the complementary use of other evaluation approaches such as *ex ante* simulations, *mixed method analysis* drawing on both qualitative and quantitative data, and *process evaluations,* can serve as valuable complements to impact evaluations. These approaches have many useful applications, such as to estimate the effect of reforms before they are implemented, to help focus core impact evaluation questions, to track program implementation, and to interpret the results from impact evaluations (World Bank, 1996).

Monitoring program implementation, most often through the use of administrative data, is critical in an impact evaluation.

Ex ante simulations are evaluations that use available data to simulate the expected effects of a program or policy reform on outcomes of interest.

Mixed method approaches that combine quantitative and qualitative data are a key supplement to impact evaluations based on the use of quantitative data alone, particularly to help generate hypotheses and focus research questions before quantitative data are collected and to provide perspectives and insights on a program's performance during and after program implementation.

Evaluations that integrate qualitative and quantitative analysis are characterized as using *mixed methods* (Bamberger, Rao, and Woolcock 2010).

In developing a mixed method approach, Creswell (2014) defines three basic approaches:

i. *Convergent parallel.* Both quantitative and qualitative data are collected at the same time and used to triangulate findings or to generate early results about how the program is being implemented and perceived by beneficiaries.

ii. *Explanatory sequential.* Qualitative data provide context and explanations for the quantitative results, to explore outlier cases of success and failure, and to develop systematic explanations of the program's performance as it was found in the quantitative results. In this way, qualitative work can help explain why certain results are observed in the quantitative analysis, and can be used to get inside the "black box" of what happened in the program (Bamberger, Rao, and Woolcock 2010).

iii. *Exploratory sequential.* The evaluation team can use focus groups, listings, interviews with key informants, and other qualitative approaches to develop hypotheses as to how and why the program would work, and to clarify research questions that need to be addressed in the quantitative impact evaluation work.

Process evaluations focus on how a program is implemented and operates, assessing whether it conforms to its original design and documenting its development and operation.

A process evaluation should include the following elements, often drawn from a results chain or logic model, complemented by program documents and interviews with key informants and beneficiary focus groups (World Bank, 1996):

- ✓ Program objectives and the context in which the program is operating
- ✓ Description of the process used to design and implement the program
- ✓ Description of program operations, including any changes in operations
- ✓ Basic data on program operations, including financial and coverage indicators
- ✓ Identification and description of intervening events that may have affected implementation and outcomes
- ✓ Documentation, such as concept notes, operations manuals, meeting minutes, reports.

Cost-Benefit and Cost-Effectiveness Analysis

It is critically important that impact evaluation be complemented with information on the cost of the project, program, or policy being evaluated. Once impact evaluation results are available, they can be combined with information on program costs to answer two additional questions:

i. *What is the benefit that a program delivers for a given cost?* Cost-benefit analysis estimates the total expected benefits of a program, compared to its total expected costs. It seeks to quantify all of the costs and benefits of a program in monetary terms and assesses whether benefits outweigh costs.
ii. Policy makers should assess which program or alternative is most cost effective in reaching a particular goal. How do various program implementation alternatives compare in cost-effectiveness? This cost effectiveness analysis compares the relative cost of two or more programs or program alternatives in reaching a common outcome, such as agricultural yields or student test scores.

In a cost-benefit or cost-effectiveness analysis, impact evaluation estimates the benefit or effectiveness side, and cost analysis provides the cost information.

CHAPTER V. INDICATORS FOR MONITORING & EVALUATION

5.1 CONSTRUCTING A THEORY OF CHANGE

A *theory of change* is a description of how an intervention is supposed to deliver the desired results (Reeler, 2007). It describes the causal logic of how and why a particular program, program modality, or design innovation will reach its intended outcomes. As one of the first steps in the evaluation design, constructing a theory of change can help specify the research questions.

Theories of change depict a sequence of events leading to outcomes; they explore the conditions and assumptions needed for the change to take place, make explicit the causal logic behind the program, and map the program interventions along logical causal pathways.

The best time to develop a theory of change for a program is at the beginning of the design process, when stakeholders can be brought together to develop a common vision for the program, its goals, and the path to achieving those goals. Stakeholders can then start implementing the program from a common understanding of the program, its objectives, and how it works. Program designers should also review the literature for accounts of experience with similar programs, and verify the contexts and assumptions behind the causal pathways in the theory of change they are outlining.

5.2 THE CONCEPT AND IMPORTANCE OF PERFORMANCE MONITORING INDICATORS

Performance indicators are measurements of project outputs, outcomes, impacts, and inputs that are monitored during project implementation to assess progress toward project objectives. They are also used later to evaluate a project's success. Indicators organize

information in a way that clarifies the relationships between a project's impacts outcomes, outputs, and inputs and help to identify problems along the way that can impede the achievement of project objectives.

It is important to set objectives and assessment indicators that are realistic as too often these steps are taken without consulting with the people who are the so-called 'targets' or 'primary stakeholders' of the project. Setting indicators with key stakeholders and communities is important for the following reasons (World Bank, 1996):

- The process results in more realistic, meaningful and achievable indicators than those set by
- Top-down methods
- The process can often highlight the different information needs and ideas of change of different stakeholders and community groups.
- The focus is not just on *what* is measured but on *how* it is measured and *who* has decided which indicators are important.
- Information about *why* and *how* change has happened and how important that change is to those affected is more likely to emerge.
- The process helps to increase community ownership of and involvement in projects, awareness, mutual learning and empowerment. This can increase the potential that the goals of the program have positive impacts of various kinds.

What indicators can and cannot tell us

Indicators **can tell us** things such as:

- To what extent our program objectives have been met
- What progress our project or program has made
- The extent to which our targets have been met
- That a change we are interested in is happening

Indicators **cannot tell us:**

- Why our program or project has made a difference
- Why and how change occurs
- How our communication activities should be undertaken

5.3 DEVELOPING A RESULT CHAIN

A results chain is one way of depicting a theory of change. Other approaches include theoretical models, logic models, logical frameworks, and outcome models. Each of these models includes the basic elements of a theory of change: a causal chain, a specific cation of outside conditions and influences, and key assumptions. A results chain establishes the

causal logic from the initiation of the program, beginning with resources available, to the end, looking at long-term goals. It sets out a logical outline of how a sequence of inputs, activities, and outputs for which a program is directly responsible interacts with behavior to establish pathways through which impacts are achieved (figure 4). A basic results chain will map the following elements (Zeller and Nielson, 2013):

- ✓ *Inputs.* Resources at the disposal of the project, including staff and budget.
- ✓ *Activities.* Actions taken or work performed to convert inputs into outputs.
- ✓ *Outputs.* The tangible goods and services that the project activities produce; these are directly under the control of the implementing agency.
- ✓ *Outcomes.* Results likely to be achieved once the beneficiary population uses the project outputs; these are usually achieved in the short to medium term and are usually *not* directly under the control of the implementing agency.
- ✓ *Final outcomes.* The final results achieved indicating whether project goals were met. Typically, final outcomes can be influenced by multiple factors and are achieved over a longer period of time.

The results chain covers both implementation and results. *Implementation* concerns the work delivered by the project, including inputs, activities, and outputs. *Results* consist of the outcomes and final outcomes, which are not under the direct control of the project and which are contingent on behavioral changes by program beneficiaries. *A good results chain helps surface assumptions and risks implicit in the theory of change.* Policy makers are best placed to articulate the causal logic and the assumptions on which it relies—as well as the risks that may affect the achievement of intended results. The team that conducts the evaluation should draw out these implicit assumptions and risks in consultation with policy makers. A good results chain will also reference evidence from the literature regarding the performance of similar programs.

5.4 THE LOGICAL FRAMEWORK

The ZOPP methodology can be applied at all stages of project preparation and implementation and/or applied annually in projects to update planning as needed (World Bank, 2001). ZOPP workshops typically last one week, though can last just one day or as long as two weeks. The log frame/ZOPP approach is a tool to create project objectives. Project objectives need to be structured to match the analysis of problems that the project is trying to overcome. Problem analysis is a brainstorming technique by which stakeholders identify the causes and effects of problems. Project objectives are structured to resolve those problems and can be represented as a mirror image of the problem tree diagram.

The output of the ZOPP approach is a project planning matrix, summarizing and structuring the main elements of a project as well as highlighting logical linkages between intended inputs, planned activities and expected results (GTZ, 1997).

Table 2. Project planning matrix

Strategy	Indicators	Assumptions	Indicators of the assumptions
Overall goal: superior strategic goal of the project (the project's contribution to policy or program objectives)	How the overall objective will be measured including quantity, quality and time		
Development goal: the changed situation designed by the target groups	How to recognize whether the development goal has been achieved		
Project purpose: change in actions of the users of the project's services	How to recognize and measure that project purpose has been achieved including quantity, quality and time	Matters outside the influence of the target groups which must happen for them to achieve their development goal	How to recognize that the assumption has taken place
Results: products and services generated by the project management	How the results are to be measured including quantity, quality and time	Matters outside the project which must happen if results are to be achieved (assumptions that must hold true to deliver the results)	How to recognize that the assumption has taken place
Activities to achieve the results	Quantities and costs		

Source: GTZ (1997, 2007)

5.5 TYPES OF PERFORMANCE INDICATORS

There are various types of indicators that measure different variables relevant for project implementation and evaluation. These are input indicators, process indicators, outcome indicators, impact indicators, and exogenous indicators (World Bank, 2001).

Input indicators are quantified, time-bound statements of resources to be provided. Input indicators are often left out of discussions of project monitoring, though they are part of the management information system. Examples of input indicators include:

- vehicle operating costs
- levels of financial contributions from the government or co-financiers
- appointments of staff
- provision of buildings
- status of enabling legislation

Process indicators measure what happens during implementation. Examples:

- Date by which building site clearance must be completed
- Latest date for delivery of fertilizer to distribution sites

- Number of distribution sites reporting availability of fertilizer
- Number of farmers receiving fertilizer
- Number of farmers in possession of fertilizer.

Outcome indicators show the immediate physical and financial outputs of the project, including physical quantities, organizational strengthening and initial flows of services. Examples include:

- Kilometers of all highway completed by a particular date
- Percentage of farmers attending a crop demonstration
- Number of teachers trained in textbook use
- Cost per kilometer of road construction
- Crop yield per hectare
- Time taken to process a credit application

Impact refers to medium or long-term developmental change that can be attributed *solely* to the project and policy. Measures of change often involve complex statistics about economic or social welfare and depend on data that are gathered from beneficiaries, through qualitative and/or quantitative research methods. Early indications of impact may be obtained by surveying beneficiaries about their perceptions of the project's services. This type of leading indicator has the twin benefits of consulting with stakeholders and providing advance warning of problems that might arise. Examples of impact indicators include:

- Percent decrease in area harvested, percent increase in household income
- Through sales of wood and non-wood products.
- Percent increase in yield per hectare, percent increase in household
- Income from farm production, percentage change in on-farm labor requirements

5.6 EXAMPLES OF SOCIO-ECONOMIC INDICATORS FOR TERRITORIAL DYNAMICS

The OECD activity on Territorial Statistics and Indicators (TSI) is undertaking pioneer work to establish an international statistical database on comparable sub-national territories (Report on Territorial Indicators of Socio-Economic Dynamics, 1999). This allows calculating *sets of territorial development indicators* revealing the huge variety of demographic, economic, social and environmental conditions and trends usually hidden behind national average figures.

Table 3. Indicators for territorial development

Territorial indicators shall provide information on:	
ECONOMIC structures & performance	**SOCIAL** well-being & cohesion
Productivity	Income and poverty
Growth	Employment
Investment	Education and skills
Innovation	Health and safety
Sectoral mix	Housing and community
DEMOGRAPHIC patterns and migration	**ENVIRONMENTAL** quality and amenity
Density	Topography and climate
Natural balance	Soils, water, species
Youth and ageing	Habitats and landscapes
Migration	Settlement and equipment
Households	Cultural heritage

Source: OECD (1999), RISI report on "Territorial Indicators of Socio-Economic Dynamics"

5.7 GENERAL PRINCIPLES FOR SELECTING INDICATORS

A clearly articulated results chain provides a useful map for selecting the indicators that will be measured along the chain. They include indicators used to monitor program implementation and to evaluate results. It is useful to engage program stakeholders from both the policy and research teams in selecting these indicators, to ensure that those selected are good measures of program performance. A widely used rule of thumb to ensure that the indicators used are good measures is summed up by the acronym SMART. Indicators should be the following (Doran, 1981):

- *Specific:* To measure the information required as closely as possible
- *Measurable:* To ensure that the information can be readily obtained
- *Attributable:* To ensure that each measure is linked to the project's eff orts
- *Realistic:* To ensure that the data can be obtained in a timely fashion, with reasonable frequency, and at reasonable cost
- *Targeted:* To the objective population

CHAPTER VI. MONITORING AND EVALUATING POLICIES AND PROJECTS

6.1 SPECIFYING EVALUATION QUESTIONS

A clear evaluation question is the starting point of any effective evaluation. The formulation of an evaluation question focuses the research to ensure that it is tailored to

the policy interest at hand. The focus is on the *impact*: that is, the changes *directly attributable* to a program, program modality, or design innovation.

In an impact evaluation, the evaluation question needs to be framed as a *well-defined, testable hypothesis.* The question should be able to frame the question in such a way that you can quantify the difference between the results obtained contrasting the treatment and comparison groups. The results chain can be used as a basis for formulating the hypothesis that you would like to test using the impact evaluation.

6.2 THE DESIGN OF MONITORING AND EVALUATION SYSTEMS

Although experts vary on the specific sequence of steps in building a results-based M&E system, all agree on the overall intent. For example, different experts propose four- or seven-step models. Regardless of the number of steps, the essential actions involved in building an M&E system are to (World Bank, 2004):

- Formulate outcomes and goals
- Select outcome indicators to monitor
- Gather baseline information on the current condition
- Set specific targets to reach and dates for reaching them
- Regularly collect data to assess whether the targets are being met
- Analyze and report the results.

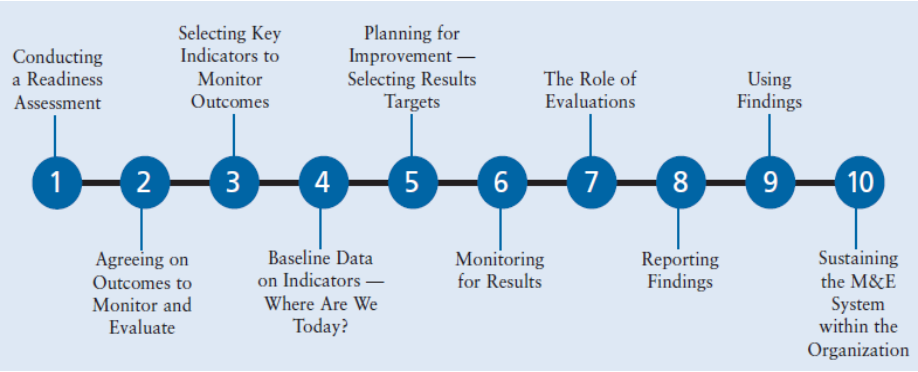

Figure 2. Ten steps to designing, building and sustaining a result-based monitoring & evaluation system
Source: The World Bank (2004): Ten steps to a results-based monitoring and evaluation system

In order to construct a M&E system must begin with the foundation of a readiness assessment. Without an understanding of the foundation, moving forward may be fraught with difficulties and, ultimately, failure. In addition, in order to be successful a ten-step result-based monitoring and evaluation system is needed.

6.3 DATA SOURCES FOR MONITORING AND EVALUATION, COLLECTION AND MANAGEMENT

The following data sources can be included in the M&E system:

Project field records: information on input and process indicators comes from project management records originating from field sites. The quality of record keeping in the field the standard for all further use of data and merits careful attention.

Surveys and studies: measuring output and impact may require the collection of data from sample surveys or other study methods (including participatory methods, if appropriate).

Data comparability: some desired indicators of impact may involve comparisons with the situation before the project or with areas not covered by the project.

On the table below are presented the main data collection instruments for impact evaluation:

Table 4. Data collection instruments

Technique	Definition and use	Strengths	Weaknesses
Case studies	Collecting information that can be descriptive or explanatory and can serve to answer the questions of how and why	Can deal with a full variety of evidence from documents, interview, observation	Good case studies are difficult to do Require specialized research and writing skills Findings are not generalizable to the population Time consuming Difficult to replicate
Focus groups	Holding focused discussions with members of target population who are familiar with pertinent issues. The purpose is to compare the beneficiaries' perspectives with abstract concepts in the evaluation's objectives.	Similar advantages to interviews (below).	Can be expensive and time consuming Must be sensitive to mixings of hierarchal levels of participants Not generalizable
Interviews	The interviewer asks questions to one or more persons & records the respondent's answers. Interviews may be formal or informal, face-to-face or not, or closed or open- ended	People & institutions can explain their experiences in their own words & settings. Flexible to allow the interviewer to pursue unanticipated lines of inquiry and to probe into issues in depth. Particularly useful where language difficulties are anticipated	Time consuming Can be expensive If not done properly, the interviewer can influence interviewees' responses

		Greater likelihood of getting input from senior officials	
Observation	Observing and recording situation in a log/diary. This includes: who is involved; what happens; when, where and how events occur. Observations can be direct (observer watches and records) or participatory (observer becomes part of the setting for a period of time)	Provides descriptive information on context and observed changes	Quality and usefulness of data is highly dependent on the observer's observational and writing skills Findings can be open to interpretation Does not easily apply within a short period of time
Questionnaires	Developing a set of survey questions whose answers can be coded consistently	Can reach a wide sample, simultaneously Allow respondents time to think before they answer Can be answered anonymously Impose uniformity by asking all respondents the same things Make data compilation and comparison easier	The quality of responses is highly dependent on the clarity of questions Sometimes it is difficult to persuade people to complete and return questionnaires Can involve forcing institutional activities & people's experiences into predetermined categories
Written document analysis	Reviewing documents such as records, administrative databases, training materials and correspondence	Can identify issues to investigate further and provide evidence of action, change and impact to support respondents' perceptions Can be inexpensive	Can be time consuming

Source: Adapted from Taschereau (1998) in Baker (2000)

6.4 INSTITUTIONAL ARRANGEMENTS AND CAPACITY BUILDING

The M&E capacity requirements of the project should be designed in the context of the evaluation capacity needs of sectoral and national institutions in the country concerned. Virtually all implementing agencies will have existing reporting systems. The M&E design should aim to build on these arrangements and to develop further the technical skills required to plan information needs, design data collection, execute studies and surveys,

analyze the data, and report results in a format that is relevant to decision makers. Table 5, shows tradeoffs between utilizing internal and external evaluators for M&E.

Table 5. Different evaluation bodies for evaluation

Evaluation body	Advantages	Disadvantages
Internal evaluator (someone involved in the project)	Knows organization, its programs and operations Understands and can interpret personal behavior and attitudes Is known to staff, so may pose no threat of anxiety or disruption Has greater chance of adopting or following-up recommendations May provide more opportunities to build national evaluation capability	Finds it harder to be objective May avoid looking at negative factors or forming negative conclusions Tends to accept organization's Assumptions Is usually too busy to participate fully Is part of the authority structure and may be constrained by organizational role conflict May not be trained in evaluation Methods May not have technical expertise
External evaluator (someone not involved in the project)	Not personally involved, so finds it easier to be objective May be free from organizational bias Can bring fresh perspective and insight May have broader experience and more experience in evaluation More easily hired for intense work Can serve as a facilitator between parties Can bring the organization into contact with additional resources	May not know the organization, its policies, procedures and personalities May be ignorant of constraints affecting feasibility of recommendations May be perceived as an adversary arousing unnecessary anxiety May be expensive (unless contracted locally) Requires more time to hire Cannot follow up on recommendations May be unfamiliar with the local political, cultural and economic environment (unless contracted locally)

Source: UNICEF, 1991

The uses of the information gained from evaluation can be structured and scheduled according to the needs of the participants. Outputs are unlikely to be measurable at less

than three-monthly intervals, and some may need much longer. Consultations with beneficiaries or surveys on their satisfaction with project services should be timed to supply information to use in planning project activities. The time period for reporting may vary with the level of management, such as monthly reporting at the district level and quarterly reporting at the region or state level. When discussing institutional arrangements, it is important to highlight the advantages and disadvantages of working with government agencies versus NGOs or firms as partners. Governments offer several advantages over NGOs: they allow for a much wider geographic scope, the results may be more likely to feed into the policy process, and there will be less concern about whether the results are dependent on a particular organizational culture. On the other hand, NGO and firms offer a much more flexible environment where it can be easier for researchers to monitor the implementation of their research design.

6.5 INTEGRATION OF QUANTITATIVE AND QUALITATIVE METHODS

Although there is an extensive literature on quantitative versus qualitative methods in impact evaluation, there is also a growing acceptance on the need to integrate these two approaches. *Quantitative data* are numerical measurements and are commonly associated with scales or metrics. *Qualitative data* are expressed not in numbers, but rather by means of language or sometimes images.

There are significant tradeoffs in selecting one technique over another. A comparison or quantitative and qualitative approaches for evaluation is provided in Table 6.

Table 6. A comparison of qualitative and quantitative approaches

	Quantitative approach	**Qualitative approach**
Objectives	To assess causality and reach conclusions that can be generalized	To understand processes, behaviors, and conditions as perceived by the groups or individuals being studied
Data collection instrument	Structured, formal, pre designed questionnaires	In-depth, open-ended interviews, direct observation, written documents such as open-ended written items on questionnaires, personal diaries, and program records
Sampling	Probability sampling	Purposive sampling
Methodology for analysis	Predominantly statistical analysis	Triangulation, systematic content analysis, and graduation aggregation of data based on selected themes

Source: Prennushi et al. (2000)

Integrating quantitative and qualitative evaluations can often be the best vehicle for meeting the project's information needs. In combining the two approaches, qualitative methods can inform the key impact evaluation questions, survey the questionnaire or the

stratification of the quantitative sample, and analyze the social, economic, and political context within which a project takes place, while quantitative methods can inform qualitative data collection strategies, design a statistically representative sample to inform the extent to which the results observed in the qualitative work can be generalized to a larger population.

6.6 COST BENEFIT AND COST EFFECTIVENESS ANALYSIS ON IMPACT EVALUATIONS

It is critically important that impact evaluation becomes complemented with information on the cost of the project, program, or policy being evaluated. Once impact evaluation results are available, they can be combined with information on program costs to answer two additional questions. *Cost-benefit analysis* estimates the total expected benefits of a program, compared to its total expected costs. It seeks to quantify all of the costs and benefits of a program in monetary terms and assesses whether benefits outweigh costs. The *cost-effectiveness analysis* compares the relative cost of two or more programs or program alternatives in reaching a common outcome, such as agricultural yields or student test scores. In a cost-benefit or cost-effectiveness analysis, impact evaluation estimates the benefit or effectiveness side, and cost analysis provides the cost information.

6.7 SAMPLE CHECKLIST OF ACTIVITIES AND ACTIONS FOR EFFECTIVE M&E

Impact evaluations are complex undertakings with many moving parts. The following checklist highlights the core elements of a well-designed impact evaluation (Gertler et al. 2016):

- A concrete and relevant policy question—grounded in a theory of change—that can be answered with an impact evaluation
- A robust methodology, derived from the operational rules of the program, to estimate a counterfactual that shows the causal relationship between the program and outcomes of interest
- A well-formed evaluation team that functions as a partnership between a policy team and a research team
- A respect for ethical standards and consideration of human subjects in the design and implementation of the evaluation and related data collection, as well as attention to open science principles to ensure transparency
- A sample with sufficient statistical power to allow policy-relevant impacts to be detected
- A methodology and sample that provide results generalizable for the population of interest

- High-quality data that provide the appropriate information required for the impact evaluation, including data for the treatment and comparison groups, data at baseline and follow-up, and information on program implementation and costs
- An engagement strategy to inform policy dialogue through the implementation of the impact evaluation, as well as an impact evaluation report and associated policy briefs disseminated to key audiences in a timely manner.

CHAPTER VII. METHODS FOR MONITORING, EVALUATION & IMPLEMENTATION

7.1 SAMPLING AND POWER CALCULATIONS

Once you have chosen a method to select a comparison group and estimate the counterfactual, one of the next steps in undertaking an impact evaluation is to determine the data that will be needed and the sample required to estimate differences in outcomes between the treatment group and the comparison group. Sampling and power calculations require specific technical skills and are often commissioned to a dedicated expert. In this section, will be described the basics of performing sampling and power calculations, and we highlight the elements that the policy team needs to be able to provide to technical experts.

7.2 DRAWING A SAMPLE

Sampling is the process of drawing units from a population of interest to estimate the characteristics of that population. Sampling is often necessary, as typically it is not possible to directly observe and measure outcomes for the entire population of interest. The process by which a sample is drawn from the population of interest is crucial. The principles of sampling provide guidance to draw representative samples. In practice, there are three main steps to draw a sample:

i. The ***population of interest*** needs to be clearly defined. This requires accurately specifying the unit within the population of interest for which outcomes will be measured, and clearly defining the geographic coverage or any other relevant attributes that characterize the population of interest (World Bank, 2012).

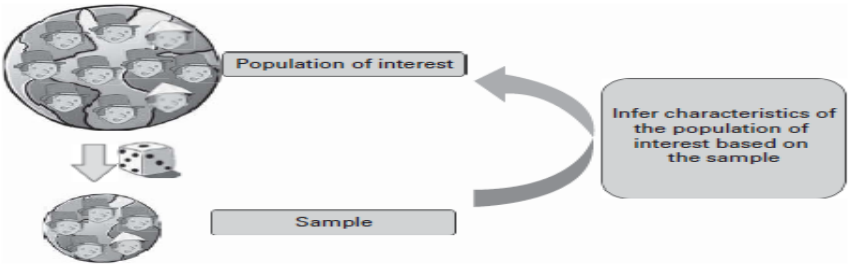

Figure 3. Using a sample to infer average characteristics of the population of interest
Source: The World Bank Group (2012): Impact evaluation in practice

ii. Second, once the population of interest has been defined, a ***sampling frame*** must be established. The sampling frame is the most comprehensive list that can be obtained of units in the population of interest. An adequate sampling frame is required to ensure that the conclusions reached from analyzing a sample can be generalized to the entire population.

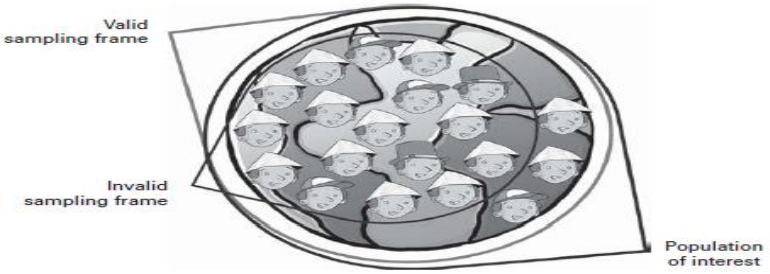

Figure 4. A valid sampling frames
Source: The World Bank Group (2012): Impact evaluation in practice

iii. ***Probabilistic sampling*** methods are the most rigorous, as they assign a well-defined probability for each unit to be drawn. The three main probabilistic sampling methods are the following:

- ***Random sampling***. Every unit in the population has exactly the same probability of being drawn.
- ***Stratified random sampling*** is performed when the population is divided in two groups (for example, male and female), and this method is used within each group. As a result, every unit in each group has the same probability of being drawn.
- ***Cluster sampling***. Units are grouped in clusters, and a random sample of clusters is drawn. This means that each cluster has a well-defined probability of being selected, and units within a selected cluster also have a well-defined probability of being drawn.

Non-probabilistic sampling can create serious sampling errors. When such a non-probabilistic sampling procedure is used, it is likely that the sample will not be representative of the population of interest as a whole.

7.3 DECIDING ON THE SIZE OF A SAMPLE FOR IMPACT EVALUATION

Sampling describes the process of drawing a sample of units from a population of interest to estimate the characteristics of that population. The calculations to determine how large the sample must be are called power calculations. The basic intuition behind power calculations by focusing on the simplest case is described as: an evaluation conducted using a randomized assignment method, testing the effectiveness of a program against a comparison group that does not receive an intervention, and assuming that noncompliance is not an issue.

7.3.1 Power calculations

Power calculations indicate the minimum sample size needed to conduct an impact evaluation and to convincingly answer the policy question of interest. In particular, power calculations can be used to (Zeller and Nielson, 2013):

- *Assess whether existing data sets are large enough to conduct an impact evaluation.*
- *Avoid collecting too little data.* If the sample is too small, it may be difficult to detect positive impact, even if it existed, and may thus conclude that the program had no effect.
- *Help make decisions about adequate sample size.* Larger sample sizes provide more accurate estimates of program impacts, but collecting information can be very costly.

7.3.2 Estimating average outcomes and comparison groups

Assume that we are interested in estimating the impact of a nutrition program on the weight of children at age two, and that 200,000 children are eligible for the program. From all eligible children, 100,000 were randomly assigned to participate in the program. The 100,000 eligible children who were not randomly assigned to the program serve as the comparison group. As a first step, we will need to estimate the average weight of the children who participated and the average weight of those who did not.

To determine the average weight of participating children, one could weigh every one of the 100,000 participating children and then average the weights. Doing that would be extremely costly. The average can be estimated using the average weight of a sample drawn from the population of participating children. When a sample is small, the average weight

constitutes a very imprecise estimate of the average in the population. In general, the more observations in the sample, the more precise the statistics obtained from the sample will be.

The same will be true for nonparticipating children: as the sample of nonparticipating children gets larger, we will know more precisely what that population looks like.

7.4 TYPES OF DATA THAT ARE NEEDED

The type of data that are needed in order to complete a good evaluation are as follow (Zeller and Nielson, 2013):

- *Data about outcomes*. Data on outcome indicators that the program indirectly affects, or indicators that capture unintended program effects, will maximize the value of the information that the impact evaluation generates, as well as the understanding of the program's overall effectiveness.
- *Data about intermediate outcomes*. Data on intermediary outcomes are useful to help understand the channels through which the program evaluated has impacted, or has not impacted, the final outcomes of interest.
- *Data about program activities and outputs*. Indicators are also required for the part of the results chain that describes program activities and outputs. In particular, *program monitoring data* can provide essential information about the delivery of the intervention, including who the beneficiaries are and which program benefits or outputs they may have received.
- *Additional data.* Other data required by the impact evaluation can depend on the methodology used. Data on other factors that may affect the outcome of interest may be needed to control for outside influences.

7.5 USING EXISTING QUANTITATIVE DATA

To determine whether existing data can be used in a given impact evaluation, a range of questions must be considered (Prennushi et al. 2000):

- *Sampling*. Are existing data available for both the treatment and comparison groups? Are existing samples drawn from a sampling frame that coincides with the population of interest? Were units drawn from the sampling frame based on a probabilistic sampling procedure?
- *Sample size*. Are existing data sets large enough to detect changes in the outcome indicators with sufficient power? The answer to this question depends on the choice of the outcome indicators, as well as on the results of the power calculations.

- *Availability of baseline data.* Are the existing data available for both the treatment and comparison groups prior to the rollout of the program or innovation to be evaluated? The availability of baseline data is important to document balance in preprogram characteristics between treatment and comparison groups when randomized methods are used, and critical for the implementation of quasi-experimental designs.
- *Frequency*. Are the existing data collected frequently enough? Are they available for all units in the sample over time, including for the times when the outcome indicators need to be measured according to the results chain and the logic of the intervention?
- *Scope*. Do existing data contain all the indicators needed to answer the policy questions of interest, including the main outcome indicators and the intermediary outcomes of interest?
- *Linkages to program monitoring information*. Can existing data be linked to monitoring data on program implementation, including to observe which units are in the treatment and comparison groups, and whether all units assigned to the treatment group received the same benefits?
- *Unique identifiers.* Do unique identifiers exist to link across data sources?

7.6 COLLECTION OF NEW SURVEY DATA

The collection of new data provides the flexibility to ensure that all the necessary indicators are measured for a comprehensive assessment of program performance.

Most impact evaluations require survey data to be collected, including at least a *baseline survey* before the intervention or innovation to be evaluated, and a *follow-up survey* after it has been implemented. Survey data may be of various types, depending on the program to be evaluated and the unit of analysis. Once it is decided to collect survey data for the evaluation, it will be needed to:

- Determine who will collect the data
- Develop and pilot the data collection instrument
- Conduct fieldwork and undertake quality control
- Process and store the data

CHAPTER VIII. METHODS FOR IMPACT ASSESSMENT OF POLICIES

8.1 CASUAL INFERENCE AND COUNTERFACTUALS

8.1.1 Casual inference

Many policy questions involve cause-and-effect relationships. Impact evaluations pursue to answer such cause-and-effect questions precisely. Although cause-and-effect questions are common, answering them is not always easy. Impact evaluations help us overcome the challenge of establishing causality by empirically establishing to what extent a particular program *and that program alone* contributed to the change in an outcome (Pearl, 2009). The answer to the basic impact evaluation question, on the impact or causal effect of a program (P) on an outcome of interest (Y) is given by the basic impact evaluation formula:

$$\boldsymbol{\Delta = (Y \mid P = 1) - (Y \mid P = 0)}$$

This formula states that the causal impact of a program (P) on an outcome (Y) is the difference between the outcome (Y) *with* the program (in other words, when $P = 1$) and the same outcome (Y) *without* the program (that is, when $P = 0$).

Income, then the causal impact of the vocational training program is the difference between a person's income (Y) after participation in the vocational training program (in other words, when $P = 1$) and the same person's income (Y) at the same point in time if he or she had not participated in the program (in other words, when $P = 0$).

The basic impact evaluation formula is valid for any unit that is being analyzed—a person, a household, a community, a business, a school, a hospital, or other unit of observation that may receive or be affected by a program.

8.1.2 Counterfactuals

At any given moment in time, a unit either participated in the program or did not participate. The unit cannot be observed simultaneously in two different states (in other words, with and without the program). This is called the *counterfactual problem*: How do we measure what would have happened if the other circumstance had prevailed? Although we can observe and measure the outcome (Y) for a program participant ($Y \mid P = 1$), there are no data to establish what her outcome would have been in the absence of the program ($Y \mid P = 0$). In the basic impact evaluation formula, *the term* ($Y \mid P = 0$) *represents the counterfactual* (Morgan and Winship, 2007). For example, let us assume that Farmer X that develops his activity in Albania, receives 10,000 euros, and we want to measure the impact of the financial direct payments on the production of apples for this farmer. If we

consider a perfect clone for Farmer X, the evaluation would be easy: we could just compare the amount of the increased production of Farmer X (6 units) after receiving the financial payment, with the amount of production increased by his clone, let's say Farmer Y, (4 units), who doesn't receive financial payment because of a problem in terms of documentation.

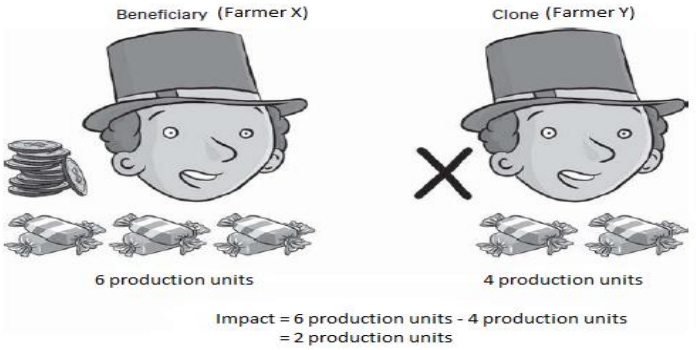

Figure 5. An example of counterfactual estimation
Source: The World Bank Group (2012): Impact evaluation in practice

In this case, the impact of direct payments on the production units of apples would be two: the difference between the number of production units increased due to the impact from subsidies (6 units) and the number of production units increased without benefiting subsidies in the form of financial payments, (4 units).

Estimating the Counterfactuals

The key to estimating the counterfactual for program participants is to move from the individual or unit level to the group level. Although no perfect clone exists for a single unit, we can rely on statistical properties to generate two *groups* of units that, if their numbers are large enough, are statistically indistinguishable from each other at the group level. The group that participates in the program is known as the *treatment group*, and its outcome is $(Y \mid P = 1)$ after it has participated in the program. The statistically identical *comparison group* (sometimes called the *control group*) is the group that remains unaffected by the program, and allows us to estimate the counterfactual outcome $(Y \mid P = 0)$: that is, the outcome that would have prevailed for the treatment group had it not received the program.

Methods for constructing comparison groups

There are two common, but highly risky, methods of constructing comparison groups that many times lead to inappropriate ("counterfeit") estimates of the counterfactual (Morgan and Winship, 2007):

- *Before-and-after comparisons* (also known as *pre-post* or *reflexive comparisons*) compare the outcomes of the same group before and after participating in a program.

- *Enrolled-and-non enrolled* (or *self-selected*) *comparisons* compare the outcomes of a group that chooses to participate in a program with those of a group that chooses not to participate.

8.2 RANDOMIZED ASSIGNMENT

This method not only provides program administrators with a fair and transparent rule for allocating scarce resources *among equally deserving populations*, but also represents the strongest method for evaluating the impact of a program.

8.2.1 Randomized assignment of treatment

Randomized assignment of treatment is considered the gold standard of impact evaluation. Under randomized assignment, every eligible unit (for example, an individual, household, business, school, hospital, or community) has the same probability of being selected for treatment by a program.

Why randomized assignment is a fair and transparent way to assign scarce program resources? Randomized assignment can often be derived from a program's operational rules. The program could be assigned on a first-come, first-served basis, or based on observed characteristics (for example, serving the poorest areas first); or selection could be based on unobserved characteristics (for example, letting individuals sign up based on their own motivation and knowledge) or on a lottery.

8.2.2 When can randomized assignment be used?

Randomized assignment can be used as a program allocation rule in one of two specific scenarios:

1. *When the eligible population is greater than the number of program spaces available.*
2. *When a program needs to be gradually phased in until it covers the entire eligible population*.

8.3 INSTRUMENTAL VARIABLES

A method called *instrumental variables* (IV) can help us evaluate programs with imperfect compliance, voluntary enrollment, or universal coverage. Generally, to estimate impacts, the IV method relies on some external source of variation to determine treatment status. The method has wide ranging applications beyond impact evaluation.

8.3.1 Types of impact estimates

An impact evaluation always estimates the impact of a program by comparing the outcomes for a treatment group with the estimate of the counterfactual obtained from a

comparison group. When we analyzed the randomized assignment, we assumed *full compliance* with treatment: that is, all units to whom a program has been offered actually enroll, and none of the comparison units receive the program. In this scenario, we estimate the *average treatment effect* (ATE) for the population.

In practice, programs typically offer the opportunity of treatment to a particular group, and some units participate while others do not. In this case, without full compliance, impact evaluations can estimate the effect of *offering* a program or the effect of *participating* in the program.

In the absence of full compliance in the treatment group, the estimated impact Δ is called the *intention-to-treat* (ITT) when comparing groups to which the program has randomly been *offered* (in the treatment group) or not (in the comparison group), regardless of whether or not those in the treatment group actually enroll in the program. By contrast, we might also be interested in knowing the impact of a program for the group of individuals who are offered the program and actually participate. This estimated impact is called the *treatment-on- the-treated* (TOT). The ITT and TOT will be the same when there is full compliance.

8.3.2 Imperfect compliance

In real-world social programs, full compliance with a program's selection criteria is desirable, and policy makers and evaluation teams alike usually strive to come as close to that ideal as possible.

The first case of imperfect compliance occurs when some units assigned to the treatment group choose not to enroll or are otherwise left untreated. In the teacher-training example, some teachers assigned to the treatment group do not actually show up on the first day of the course. *The second case of imperfect compliance* is when individuals assigned to the comparison group manage to participate in the program. Under these circumstances of noncompliance, a second option is to estimate what is known as the *local average treatment effect* (LATE). LATE needs to be interpreted carefully, as it represents program effects for only a specific subgroup of the population. In particular, when there is noncompliance in both the treatment group and in the comparison group, the LATE is the impact on the subgroup of compliers.

8.3.3 Randomized assignment of a program and final take-up

The program is randomly assigned at the individual level. The treatment group is assigned to the program, while the comparison group is not. Most likely, three types of individuals in the population will be identified (World Bank, 2012):

- *Enroll-if-assigned*. These are the individuals who comply with their assignment.

- *Never*. These are the individuals who never enroll in or take up the program, even if they are assigned to the treatment group.
- *Always*. These are the individuals who will find a way to enroll in the program or take it up, even if they are assigned to the comparison group.

Figure 6 presents the randomized assignment of the program and the final enrollment, when *Enroll-if-assigned*, *Never*, and *Always* types are present.

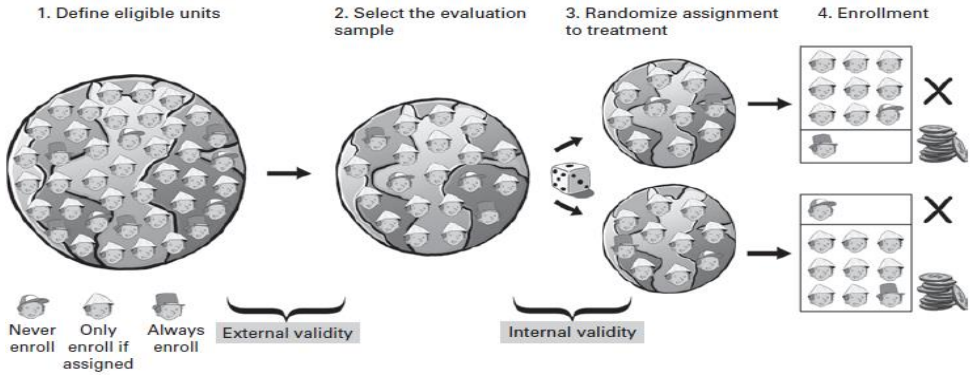

Figure 6. Randomized assignment with imperfect compliance
Source: The World Bank Group (2012): Impact evaluation in practice

8.4 REGRESSION DISCONTINUITY DESIGN

Regression discontinuity design (RDD) is an impact evaluation method that can be used for programs that have a continuous eligibility index with a clearly defined eligibility threshold (cutoff score) to determine who is eligible and who is not (Chaplin, Cook, Zurovac, Coopersmith, Finucane, Vollmer, and Morris, 2018). The RDD estimates impact around the eligibility cutoff as the difference between the average outcome for units on the treated side of the eligibility cutoff and the average outcome of units on the untreated (comparison) side of the cutoff.

To apply a regression discontinuity design, the following main conditions must be met:

1. The index must rank people or units in a continuous or "smooth" way.
2. The index must have a clearly defined cutoff score: that is, a point on the index above or below which the population is classified as eligible for the program.
3. The cutoff must be unique to the program of interest; that is, there should be no other programs, apart from the program to be evaluated, that uses the same cutoff score.
4. The score of a particular individual or unit cannot be manipulated by enumerators, potential beneficiaries, program administrators, or politicians.

8.5 DIFFERENCE IN DIFFERENCES

The ***difference in differences*** *method* compares the *changes* in outcomes over time between a population that is enrolled in a program (the treatment group) and a population that is not (the comparison group). Take, for example, a road repair program that is carried out at the district level but cannot be randomly assigned between districts and is also not assigned based on an index with a clearly defined cutoff that would permit a regression discontinuity design. District boards can decide to enroll or not enroll in the program. One of the program's objectives is to improve access of the population to labor markets, and one of the outcome indicators is the employment rate. The difference in the before and after outcomes for the enrolled group (*the first difference),* controls for factors that are constant over time in that group, since we are comparing the same group to itself. But we are still left with the factors that vary over time (*time-varying factors*) for this group. One way to capture those time-varying factors is to measure the before-and-after change in outcomes for a group that did not enroll in the program but was exposed to the same set of environmental conditions (*the second difference)*. The difference-in-differences approach does what its name suggests. It combines the two counterfeit estimates of the counterfactual (before-and-after comparisons, and comparisons between those who choose to enroll and those who choose not to enroll) to produce a better estimate of the counterfactual.

Table 7 shows the relationship of the difference in differences method. The first row contains outcomes for the treatment group before the intervention (A) and after the intervention (B). The before-and-after comparison for the treatment group is the first difference ($B - A$). The second row contains outcomes for the comparison group before the intervention (C) and after the intervention (D), so the second difference is ($D - C$).

Table 7. Calculating the difference in differences method

	After	Before	Difference
Treatment/enrolled	B	A	$B - A$
Comparison/nonenrolled	D	C	$D - C$
Difference	$B - D$	$A - C$	DD = $(B - A) - (D - C)$

	After	Before	Difference
Treatment/enrolled	0.74	0.60	*0.14*
Comparison/nonenrolled	0.81	0.78	*0.03*
Difference	*−0.07*	*−0.18*	DD = *0.14 − 0.03 = 0.11*

Source: The World Bank Group (2012): Impact evaluation in practice

The difference in differences method computes the impact estimate as follows:

1. We calculate the difference in the outcome (Y) between the before and after situations for the treatment group ($B - A$).
2. We calculate the difference in the outcome (Y) between the before and after situations for the comparison group ($D - C$).
3. Then we calculate the difference between the difference in outcomes for the treatment group ($B - A$) and the difference for the comparison group ($D - C$), or difference-in-differences (DD) = $(B - A) - (D - C)$. This difference-in-differences is our impact estimate.

We could also compute the difference in differences the other way across:

DD impact = $(B - D) - (A - C)$ = $(0.74 - 0.81) - (0.60 - 0.78) = 0.11$.

8.5.1 Positive contribution of the difference in differences method

The difference-in-differences method helps resolve this problem to the extent that many characteristics of units or individuals can reasonably be assumed to be constant over time (or *time-invariant*). Think of *observed* characteristics, such as a person's year of birth, a region's location close to the ocean, a town's climate, or a father's level of education. Using the same reasoning, we might conclude that many *unobserved* characteristics of individuals are also more or less constant over time.

The difference in differences method compares *trends* between the treatment and comparison groups. The trend for an individual is the difference in outcome for that individual before and after the program. By subtracting the *before* outcome situation from the *after* situation, we cancel out the effect of all of the characteristics that are unique to that individual and that do not change over time.

8.6 MATCHING

Matching methods can be applied in the context of almost any program assignment rules, as long as a group exists that has not participated in the program. Matching essentially uses statistical techniques to construct an artificial comparison group. For every possible unit under treatment, it attempts to find a non-treatment unit (or set of non-treatment units) that has the most similar characteristics possible. Consider a case in which you are attempting to evaluate the impact of a job training program on income and have a data set, such as income and tax records, that contains both individuals that enrolled in the program and individuals that did not enroll. The program that we are trying to evaluate does not have any clear assignment rules that explain why some individuals enrolled in the program and others did not. In such a context, matching methods enables us to identify the

set of non-enrolled individuals that look most similar to the treated individuals, based on the characteristics that you have available in your data set.

8.7 CHOOSING AN IMPACT EVALUATION METHOD

The step of choosing an impact evaluation method include:

First, we show that the program's operational rules provide clear guidance on how to find comparison groups, and thus on which method is most appropriate for your policy context.

Second, the methods introduced in the section above have different data requirements and rely on different underlying assumptions. Some methods require stronger assumptions than others to precisely estimate the changes in outcomes caused by the intervention.

Finally, we discuss how to choose the unit of intervention. In general, we prefer choosing the smallest unit of intervention feasible within operational constraints.

8.8 MANAGING IMPACT EVALUATION

An effective partnership is critical to ensuring the technical credibility and policy impact of an evaluation.

The research and policy team partnership during the evaluation

The technical quality and policy impact of the evaluation depend on an active partnership between the research team and the policy team at each stage in the evaluation: design, implementation, analysis, and dissemination.

Design stage

- First, the policy makers need to clearly structure and convey the core research questions, the accompanying theory of change, and the core indicators of interest, and ensure that the research team has a good understanding of and respect for these elements.
- Second, the research team needs to clearly understand the program's rules of operation: namely, its available resources, eligibility criteria for selecting beneficiaries, and timing for implementation.
- Third, the research team should prepare an impact evaluation plan that contains both operational and research aspects, and share this with policy makers to ensure that the evaluation is focused on the questions of interest

Implementation stage

The policy and research teams need to work together to ensure that implementation proceeds smoothly and to troubleshoot. For example, in a randomized controlled trial, the teams need to agree on the best way to randomize in practice. In addition, during this stage,

coordination is especially important to ensure fidelity between the evaluation design and program implementation.

Analysis stage

The analysis that is carried out should correspond to what is outlined in the evaluation plan and in the more detailed pre-analysis plan. The research team should provide and discuss results with the policy team at key junctures. As early as the baseline, this should include a review of the quality of the data collected and adherence to the evaluation plan. This will help ensure that the evaluation plan envisioned in the design stage remains feasible and allow any necessary adjustments to be made.

Dissemination stage

In this stage, the policy team needs to ensure that the evaluation results reach the right people at the right time in an appropriate format.

Table 8. The general outline of an impact evaluation plan

1. Introduction
2. Description of the intervention
3. Objectives of the evaluation
3.1 Hypotheses, theory of change, results chain
3.2 Policy questions
3.3 Key outcome indicators
3.4 Risks
4. Evaluation design
5. Sampling and data
5.1 Sampling strategy
5.2 Power calculations
6. Pre-analysis plan overview
7. Data collection plan
7.1 Baseline survey
7.2 Follow-up survey(s)
8. Products to be delivered
8.1 Baseline report
8.2 Impact evaluation report
8.3 Policy brief
8.4 Fully documented data sets, design and analysis protocols
9. Dissemination plan
10. Ethical protocols on protection of human subjects
10.1 Ensuring informed consent
10.2 Obtaining approval from the Institutional Review Board (IRB)
11. Time line
12. Budget and funding
13. Composition and roles of evaluation team

Source: The World Bank Group (2012): Impact evaluation in practice

CONCLUDING REMARKS

This material provides the reader with a guidance in terms of planning and implementation of a successful evaluation of a project, program and policy, related to the management and sustainable development of rural territories. The concept of evaluation comprises a wide variety of methodologies that can be applied in order to evaluate different aspects of programs in profit and non-profit organizations. This booklet provides an analysis of monitoring and evaluation, the design of an evaluation and different combined techniques and methods.

The evaluation of projects and interventions can be very helpful in terms of different aspects (World bank, 2012):

- Reminding the management, of what the project or intervention is all about, what are the goals and objectives and how these goals should be met.
- Produce results that can be used and promoted in different regions and their communities
- Produce comparisons between different programs and decide the ones that should be applied

Despite the importance of monitoring and evaluation there are certain perceptions and issues that the evaluation of a project or program should address:

- When referring to evaluation, people perceive this process as an action or activity that generates boring conclusions because of the bulky data's that are usually generated. In this context, the value of an evaluation is violated. In fact, it is important to highlight that these perceptions were more present in the past. Nowadays, evaluation focuses more on utility, relevance and validity.
- Different stakeholders believe that the process of monitoring and evaluation serve as tools that prove the failure or the success of a project or program. In reality, monitoring and evaluation serve as sources of continuous feedback and help the implementation body of that intervention to get oriented toward success.

References

- Alexander, J. W. (1954). The basic–non basic concept of urban economic function. Economic Geography, 30, 246–261.
- Andre Torre, Frederic Wallet (2016). Regional Development in Rural Areas. Analytical Tools and Public Policies. Springer. ISBN 978-3-319-02372-4.
- Bamberger, M. (2000). Integrating Quantitative and Qualitative Methods in Development Research. Washington, D.C.: World Bank.

- Bamberger, Michael, Vijayendra Rao, and Michael Woolcock. 2010. "Using Mixed Methods in Monitoring and Evaluation: Experiences from International Development." Policy Research Working Paper 5245, World Bank, Washington, DC.
- Baur, H., Bosch, M., Krall, S., Kuby, T., Lobb-Rabe, A., Schütz P.-T., and Springer-Heinze, A. (2001). Establishing Plausibility in Impact Assessment.
- Becattini, G. (1990). The Marshallian industrial districts as a socio-economic notion. In F. Pyke, G.
- Becattini & W. Sengenberger (Eds.), Industrial districts and Inter-firm cooperation in Italy. Geneva: International Institute of Labour Studies.
- Bradford, N. (2017), "Canadian regional development policy: Flexible governance and adaptive implementation", Background paper prepared for the seminar "Multi-level governance for regional economic development: a focus on flexibility and adaptability", OECD, Paris, 23 January.
- Brusco, S. (1982). The Emilian model. Cambridge Journal of Economics, 6(2), 167–184.
- Charbit, C. and O. Romano (2017), "Governing together: An international review of contracts across levels of government for regional development", OECD Regional Development Working Papers, No. 04, OECD Publishing, Paris.
- Chaplin, Duncan D.; Cook, Thomas D.; Zurovac, Jelena; Coopersmith, Jared S.; Finucane, Mariel M.; Vollmer, Lauren N.; Morris, Rebecca E. (2018). "The Internal and External Validity of the Regression Discontinuity Design: A Meta-Analysis of 15 Within-Study Comparisons". Journal of Policy Analysis and Management. 37 (2): 403–429. doi:10.1002/pam.22051. ISSN 1520-6688.
- Cour des comptes (2014), "Les Contrats des projets Etat-Regions" (CPER), Cour des comptes
- Creswell, John W. 2014. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. Thousand Oaks, CA: SAGE Publications.
- Davezies, L. (2008). La République et ses Territoires. Paris: Editions du Seuil.
- Delgado, M., & Ramos, E. (2002). Understanding the institutional evolution of the European rural policy: A methodological approach. In Xth EAAE Congress, Zaragoza, August 28–31.
- Despommier, D. (2010). The vertical farm. New York: Thomas Dunne Books. 305p.
- Doran, G. T. (1981). "There's a S.M.A.R.T. way to write management's goals and objectives". Management Review. 70 (11): 35–36.
- Duflo, E., Glennerster, R., and Kremer, M. (2006). Using Randomization in Development Economics Research: A Toolkit. MIT Department of Economics Working Paper No. 06-36.
- Eade, D. (1997). Capacity Building: An approach to people-centred development. Oxfam, UK.
- European Commission, (2015). Synthesis of Ex Ante Evaluations of Rural Development Programmes 2014 – 2020.
- European Commission. (2004). Rural development: The impact of EU research (1998–2004). Workshop report, European Commission, Luxembourg.
- Food and Agriculture Organization of the United Nations (2015). Approaches to Capacity Development in Programming, Process and Tools:
- Fowler, A. et al. (1995). Participatory Self-Assessment of NGO Capacity: INTRAC Occasional Paper Series no. 10, by Alan Fowler, with Liz Goold and Rick James, INTRAC, UK.
- GTZ (Deutsche Gesellschaft für Technische Zusammenarbeit GmbH) (1997). ZOPP Objectivesoriented Project Planning. Eschborn, Germany.

- Imas, Linda G. M., and Ray C. Rist. 2009. The Road to Results: Designing and Conducting Effective Development Evaluations. Washington, DC: World Bank.
- In Baker, J. (2000). Evaluating the Impact of Development Projects on Poverty. A Handbook for Practitioners. The World Bank, Washington, D.C.
- INTRAC (2010). Praxis Paper 23: Monitoring and Evaluating Capacity Building,
- Isard, W. (1956). Location and space economy (Boston. MA: MIT Press).
- James, R. and Hailey, J. (2007). Capacity Building for NGOs: Making it work, INTRAC, UK.
- Johansson, B., Karlsson, C., & Stough, R. R. (2001). Theories of endogenous regional growth. Heidelberg: Springer.
- Julia N. Reimers (2018). A handbook for the promotion of the agri-food sector", GIZ 2018
- Lascoumes, P., & Le Galès, P. (2004). Gouverner par les instruments. Paris: Les Presses de Science Po.
- Mollard, A., Sauboua, E., & Hirczak, M. (Eds.). (2007). Territoires et enjeux du de´veloppement régional. Paris: Quae.
- Morgan, Stephen; Winship, Chris (2007). Counterfactuals and Causal inference. Cambridge University Press. ISBN 978-0-521-67193-4.
- Mormont, M. (1990). Who is rural? Or, how to be rural: Towards a sociology of the rural. In T. Marsden, S. Whatmore & P. Lowe (Eds.), Rural restructuring (pp. 21–44). London: Fulton.
- OECD (1999), RISI report on "Territorial Indicators of Socio Economic Dynamics"
- OECD (2006). The Challenge of Capacity Development: Working towards good practice, OECD DAC Network on Governance.
- OECD (2006b). The New Rural Paradigm: Policies and Governance, OECD, Paris.
- OECD (2007). Effective Targeting of Agricultural Policies: Best Practices for Policy Design and Implementation, OECD, Paris.
- OECD (2018). Rethinking regional development policy-making, OECD multi-level governance studies, Paris.
- OECD. (2009a). Organisation for economic co-operation and development—Region at the glance. Paris: OECD.
- Ortiz, A. and Taylor, P. (2008). Learning Purposefully in Capacity Development: Why, what and when to measure? An opinion paper prepared for IIEP, July 2008.
- Paul J. Gertler et al (2016). "impact Evaluation in Practice", Second edition.
- Pearl, Judea (2009). "Causal inference in statistics: An overview" (PDF). Statistics Surveys. 3: 96–146. doi:10.1214/09-SS057.
- Porter, M. (1998). Clusters and competition: New agendas for companies, governments, and institutions. In M. Porter (Ed.), On competition (pp. 197–287). Boston: Harvard Business School Press.
- Porter, M. E. (1985). Competitive advantage. New York: The Free Press.
- Porter, M. E. (1990). The competitive advantage of nations. New York: The Free Press.
- Prennushi, G., Gloria, R., and Kalanidhi, S. (2000). Monitoring and Evaluation in Sourcebook for Poverty Reduction Strategies. Washington, D.C.: World Bank.
- Prof. Dr. Manfred Zeller and Dr. Thea Nielsen (2013). Study lectures "Monitoring and evaluation of development projects and policies", Hohenheim University.

- Rao, Vijayendra, and Michael Woolcock. 2003. "Integrating Qualitative and Quantitative Approaches in Program Evaluation." In The Impact of Economic Policies on Poverty and Income Distribution: Evaluation Techniques and Tools, edited by F. J. Bourguignon and L. Pereira da Silva, 165–90. New York: Oxford University Press.
- Reeler, D. (2007). A Theory of Social Change and Implications for Practice, Planning, Monitoring and Evaluation.
- Reid, N., Gatrell, J. D., & Ross, P. (Eds.). (2012). Local food systems in old industrial regions. Farnham: Ashgate. 264p.
- Sack, R. (1986). Human territoriality. Its theory and history. Cambridge: Cambridge University Press.
- Solow, R. M. (2000). Growth theory: An exposition. New York: Oxford University Press.
- Steve Wiggins, 2015. Agricultural Development Policy: A contemporary agenda.
- Taschereau, S. (1998). Evaluating the Impact of Training and Institutional Development Programs, a Collaborative Approach. Economic Development Institute of the World Bank, January.
- Terrier, C. (2006). L'économie présentielle, un outil de gestion du territoire, Cahier Espaces n°90, numéro spécial Observation et
- The World Bank (2004): Ten steps to a results-based monitoring and evaluation system
- The World Bank Group (2012): Impact evaluation in practice
- Tourisme, 6 pp. Paris: Editions Espaces tourisme & loisirs.
- UNICEF (1991). A UNICEF Guide for Monitoring and Evaluation. Making a difference? Evaluation Office, New York.
- United Nations Development Programme (2019): Handbook on planning, monitoring and evaluating for development results.
- World Bank (1996). Designing Project Monitoring and Evaluation. Operations Evaluation Department. Lessons and Practices. No. 8.
- World Bank (2001). Tools: Interaction Community Planning ZOPP: Goal Oriented Project Planning in Upgrading Urban Communities A Resource for Practitioners.
- World Bank Group (2016). Impact Evaluation in Practice, Second Edition.
- World Bank statistics