Introduction to Innovation Policy for Developing Countries

Module 5 - Part 2

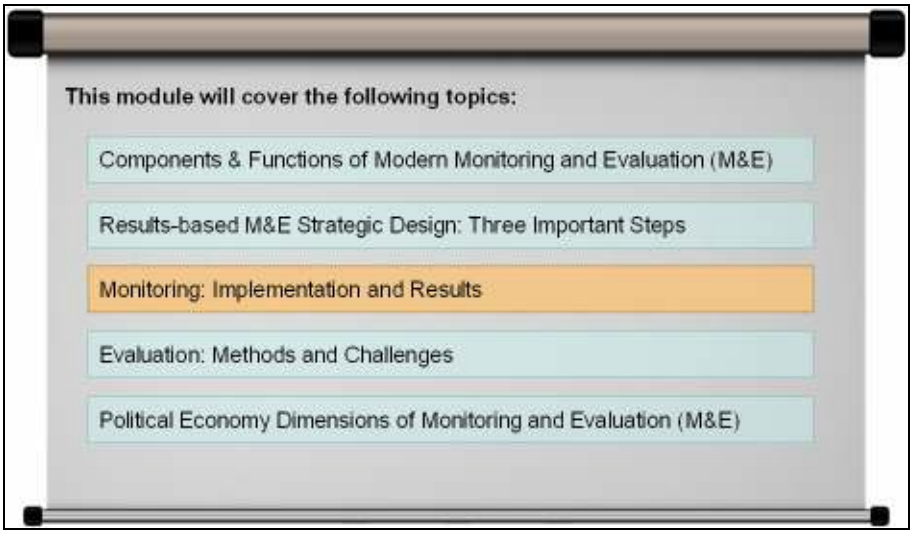
Monitoring and Evaluation (M&E) Systems

World Bank Institute

Presentation Script

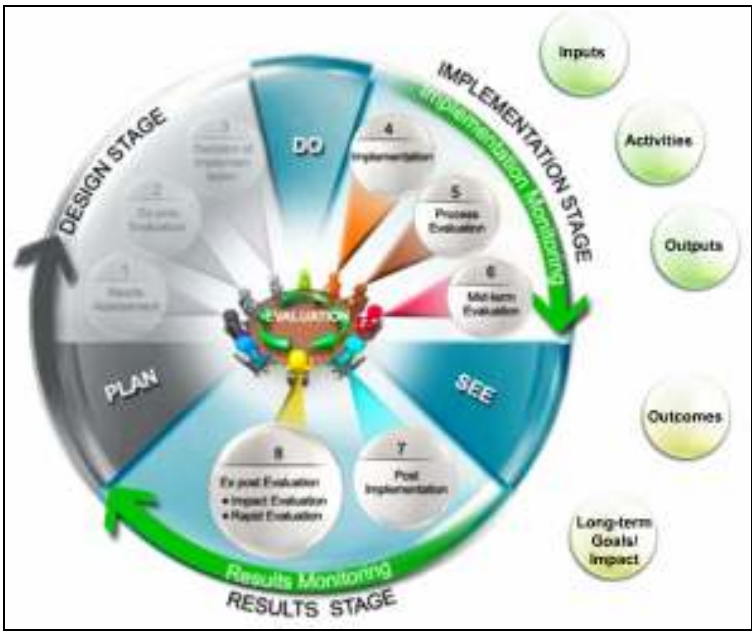
Module: Topics

And now we'll proceed to our next topic "Monitoring of Implementation and Results"



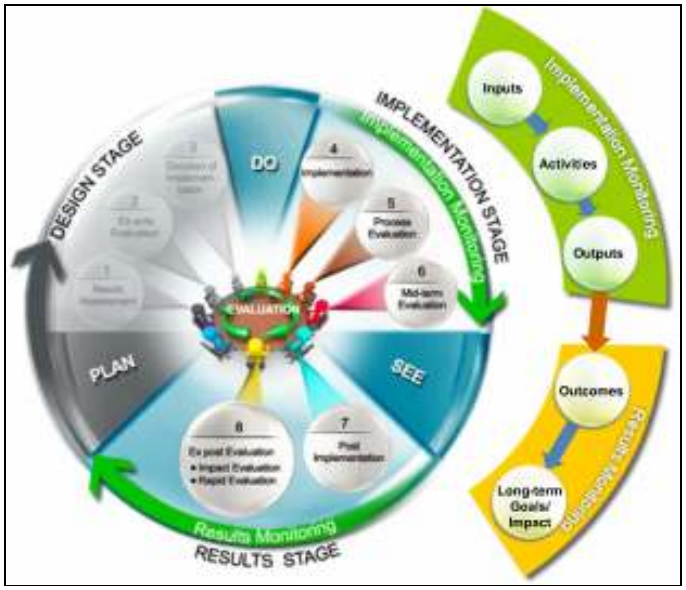
Types of Monitoring: Implementation and Results

Putting these different parts together, in order to get a complete results□based M&E framework for an area of policy, we must add that it monitors both □ Implementation (inputs, activities and outputs) and Results (Outcomes and Impact).

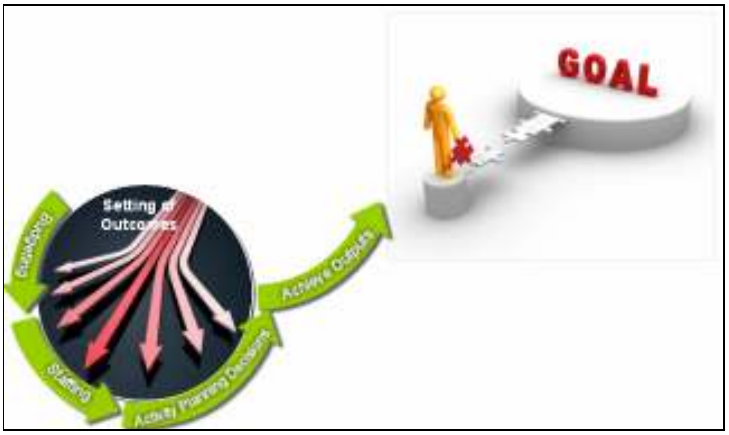


Links Between Implementation Monitoring & Results Monitoring

Implementation Monitoring tracks means and strategies (inputs, activities, and outputs), as Results Monitoring tracks Outcomes and Long-term Goals achievement, Monitoring for results in this way provides evidence on performance and alerts policymakers to changes that may be required for a given project or policy.



In theory, each element –from inputs, activities, and outputs to outcomes and goals should have its own set of indicators, baselines and targets and the causal relationships and assumptions linking each other should be identified. All elements of this framework, are derived from and informed by the setting of outcomes; and when supported by appropriate budgeting, staffing, and activity planning decisions should, in turn, enable the organization to use inputs effectively to achieve outputs and, ultimately, realise those very outcomes.



Clearly, this framework simplifies reality. Countries often struggle to integrate performance monitoring and resource allocation, and responsibility for these activities is split across organizations and agencies. Coordination and collaboration in sharing performance information is therefore essential, especially in those cases where partnerships have been established to achieve specific targets.

M&E in Operating Cycle: Collecting & Reporting

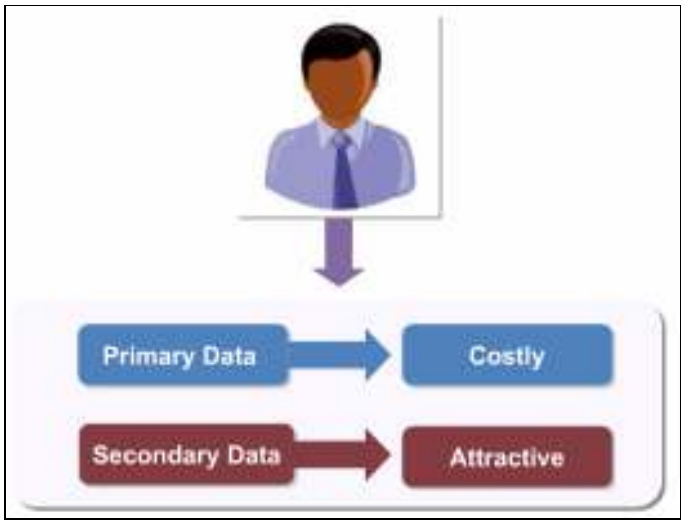
During monitoring, M&E systems a crucial task is collecting and reporting information to track whether actual results are being achieved as planned.

This process provides empirical information and helps to adjust the design and objectives as needed.

Data Collection Strategies

Data collection strategies necessarily involve some trade offs regards cost, precision, credibility, and timeliness. For example, the more rigorous methods for collecting data such as panel surveys, censuses and field experiments are more precise yet they are costly, and time consuming. If data is needed on a real□time, continuous basis to inform management decision making, less structured methods such as interviews, field visits and administrative data may be preferable, thereby surrendering precision for frequency.

Similar issues arise in the relation to data sources and the choice between primary and secondary data. Primary data –data which is collected directly by the organization itself generally provides greater precision and credibility; but sometimes it is too impractical and costly for administrators to collect primary data, as in the case of large scale household surveys, thereby making the use of secondary data more attractive.



Monitoring Results: Procedures

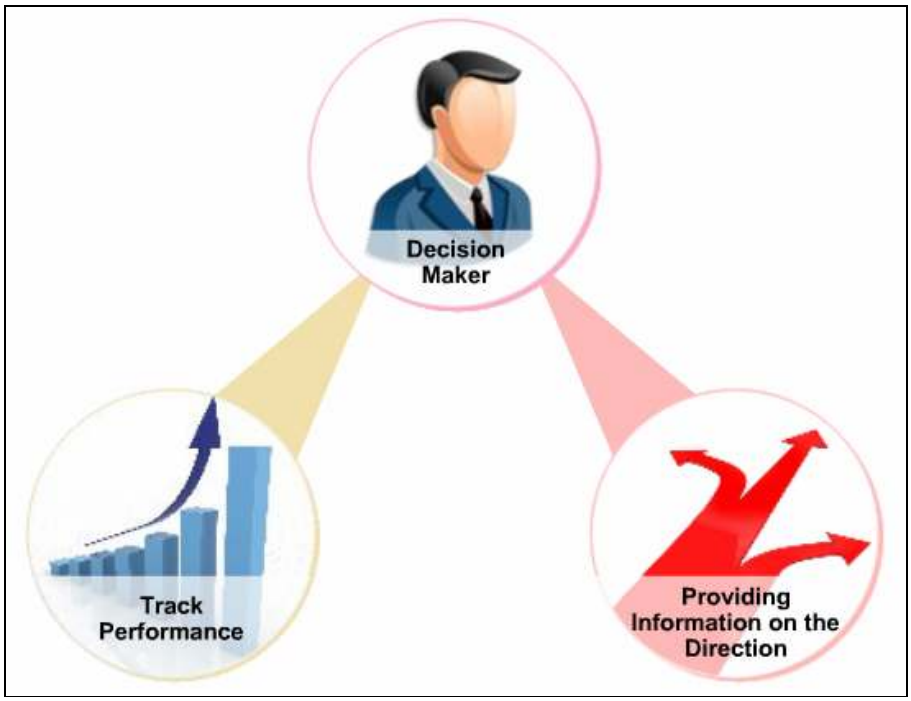
Monitoring is only as good as the data system that is in place and policymakers need to assure themselves that data sources can be accessed in a practical fashion; that data is of sufficient quality and that it is available on a regular and timely basis.

Robust data collection and analysis should have plans and procedures for each of the following areas:

- Units of analysis - for example, school district, community hospital, village, region;
- Sampling procedures;
- Data collection instruments to be used;
- Frequency of data collection;
- Expected methods of data analysis and interpretation;
- Those responsible for collecting the data;
- Data collection partners, if any;
- Those responsible for analyzing, interpreting, and reporting data;
- For whom the information is needed;
- Dissemination procedures;
- Follow-up on findings.

Challenges of "M" in M&E Innovation Policy

Constructing a monitoring system allows decision makers to continuously track performance, providing information on the direction and pace of change.



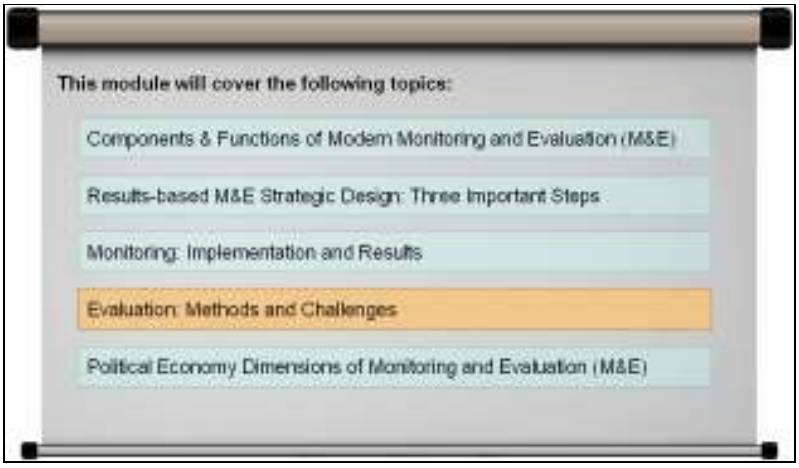
But monitoring has a number of fundamental limits: it does not provide evidence that the change has been caused by the given policy or intervention; nor does it show how changes are coming about—only that they are or are not taking place. Likewise monitoring data, in and of themselves, cannot illuminate what is good and bad about the design of the program or policy.

Fundamental limits:

- It does not provide evidence that the change has been caused by the given policy or intervention
- It does not show how changes are coming about
- It cannot illuminate what is good and bad about the design of the program or policy

Module: Topics

And now we'll proceed to our next topic "Evaluation: Methods and Challenges"



Results-based Evaluation in the Program Operating Cycle

From the diagram we see that evaluation process can happen at any moment during the operating cycle, it is period-based, and there are various kinds of evaluations that are performed specifically at different stages depending on the needs and resources.

However we'll examine methods and challenges of two types of evaluations - Ex ante (before the fact) and Ex post (after the fact) evaluation.

Challenges of "E" in M&E Innovation Policy

Evaluating the relevance, efficiency, effectiveness, impact, and sustainability of policy is not an easy task, and is complicated by a number of factors:

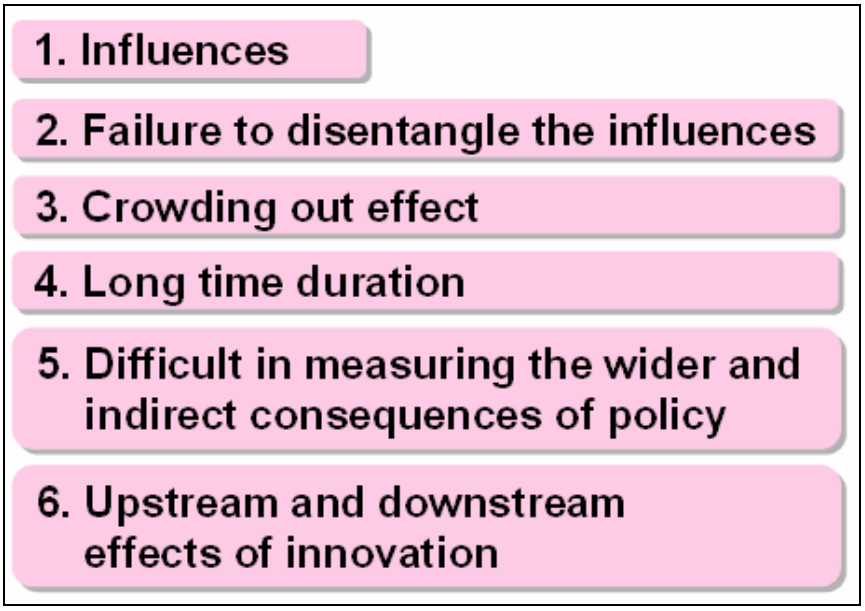
A dense web of influences may account for outcomes that policymakers observe. This is especially true of innovation where the range of policy interventions that matter –from monetary and fiscal policy, knowledge transfer, network building to framework development and IPR regimes and the factors that enable or constrain innovation activities such as entrepreneurship are seemingly infinite.

Failure to disentangle these influences from the effects of policy can lead to the problem of over-determination. By this we mean that the changes we see in innovative performance may not, or not only, be attributable to the policy itself. This makes it difficult to work out a programme's or policy's additionality, namely, the degree to which desirable outcomes would have taken place without public intervention and how the policy interacts with other policies. In fact, a crowding out effect may occur if firms that receive assistance such as public funding reduce the amount they would have invested themselves in which case support does not bring about additional activity.

Another issue is that it may take a long time before the desired impacts become apparent: for instance, the economic benefits of individual innovations often require the development of markets and only slowly generate revenues and profit – sometimes over as much as 10-15 years.

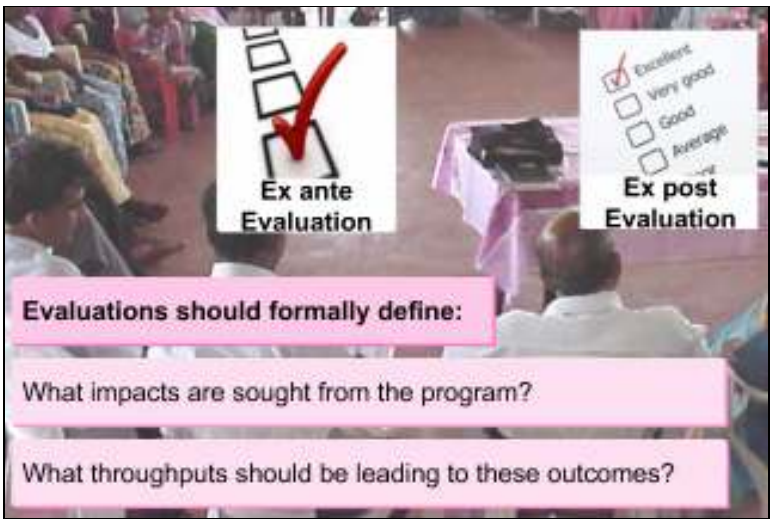
Measuring the wider and indirect consequences of policy can also be fiendishly difficult, even though these impacts are sometimes the most significant ones. There are upstream and downstream effects of innovation: a firm may demand changes in quantity and quality from its suppliers; it may open up new markets; it may change practices in its customers and it may motivate employees to spin off or set up their own firms. In turn, these may have ripple effects on other parties with whom these suppliers and customers interact to alter their behavior.

Finally, it is necessary to determine not just whether something worked but why it worked is especially important to gauge whether it would lead to similar outcomes elsewhere as well as the possibility of unintended consequences.



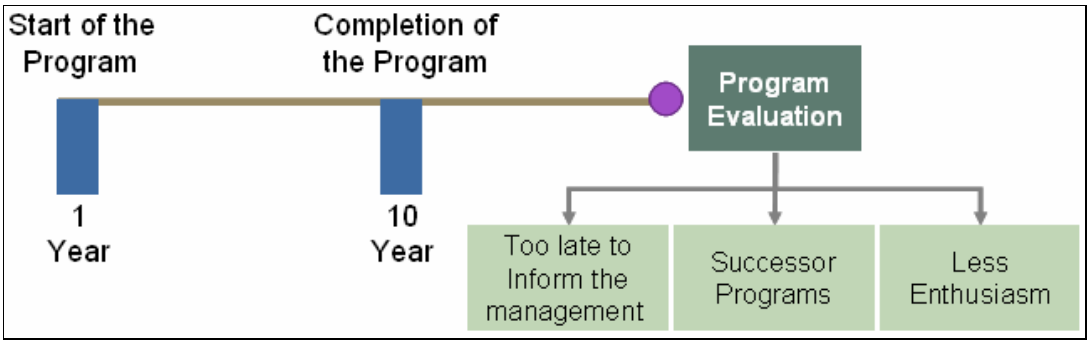
Ex ante and Ex post Evaluation

Programs should be subject to ex ante and ex post evaluation. That is, evaluations should formally define what impacts are sought from the program, and what through puts should be leading to these outcomes. Often this involves performance logic chain assessments which help flesh out the strength and logic of the causal model behind the policy or program. Here evaluation addresses the likelihood of bringing about the desired change in light of similar prior efforts and the research literature.



Results of the program should then be assessed ex post, namely, after it has been completed. Ex ante evaluation provides a valuable input to ex post evaluations, establishing a "virtual benchmark" against which program outcomes can be assessed.

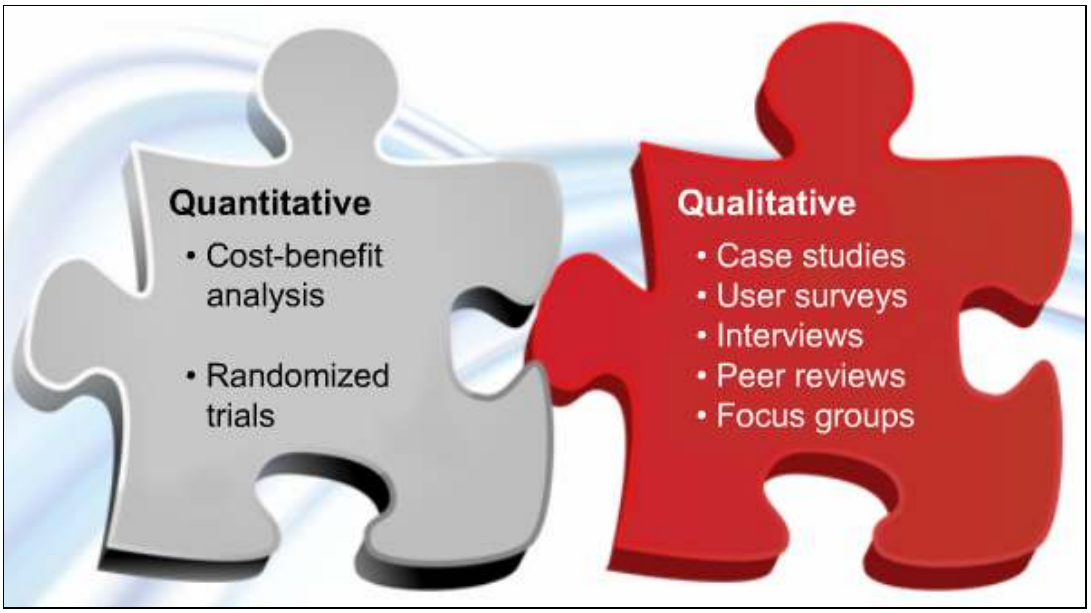
Here a critical issue is that of timing: how long after the completion of the program should the evaluation be undertaken? Ideally, evaluations should pick up lasting – rather than transitory impacts. In this way, some evaluations are carried out ten years after the program has finished. The problem is that the knowledge obtained may arrive too late to inform the management of the program or successor programs or, from a political perspective, to maintain enthusiasm for it. In this way, some real□time evaluation is often useful as a means of obtaining early warning on major lessons from the program evaluation.



Ex post Evaluation Methods

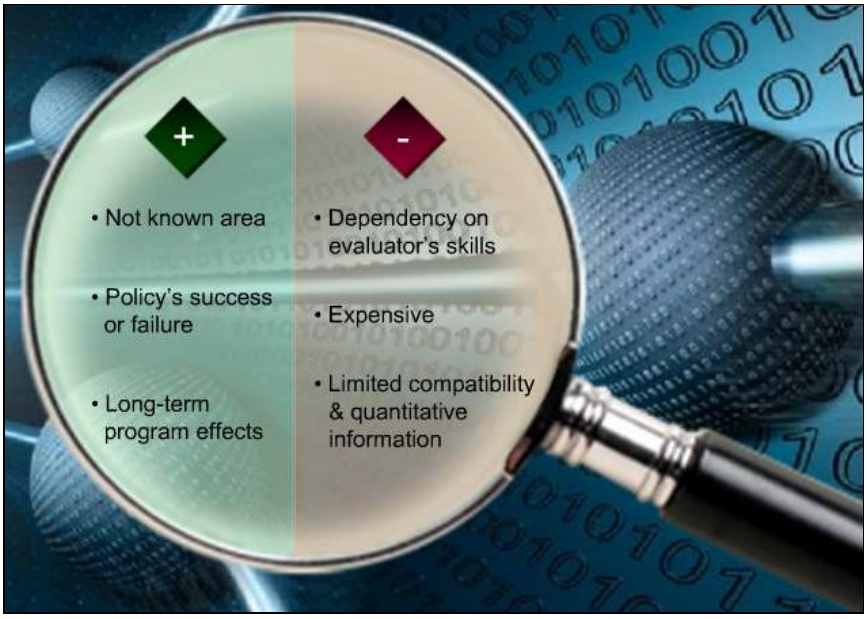
Like data collection methods, ex post evaluations draw on a variety of 'hard' and 'soft' methods, a mixture of approaches quantitative (such as cost benefit analysis and randomized trials) and qualitative (such as case studies, surveys, interviews etc) –and involve the same kind of trade offs between cost, precision, credibility, and timeliness.

In the next two slides we will examine Case Studies and User Surveys as a part of Ex post Evaluation methods.



Ex post Evaluation Methods: Case-study

Given the complexity of innovation, a particularly useful tool is the case study. Case studies favour depth over breadth providing policymakers with in depth information of what happened with the policy. They are particularly useful when little is known about an area or problem or when critical instances are involved – for instance when a policy succeeds dramatically or fails terribly. They are also suited to looking at program effects that emerge from an initiative and can yield broader insights of a condition when, over time, the results of multiple case studies are integrated. The drawbacks of case studies is that they depend heavily on evaluator's skills and experience; they are expensive when done in large numbers; they are hard to incorporate into routine monitoring and they generate limited quantitative information.

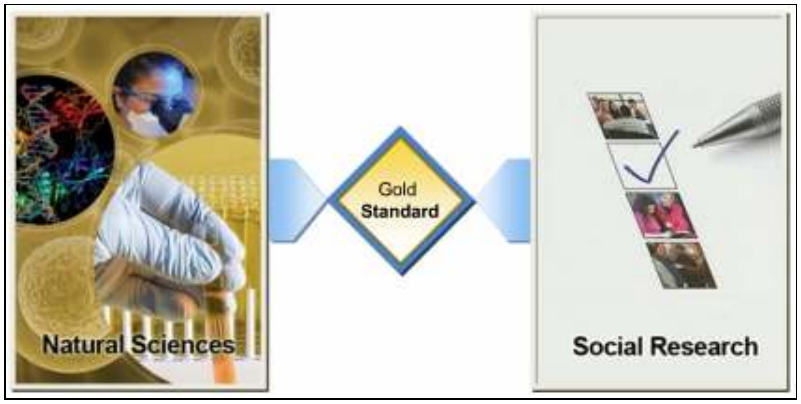


Ex post Evaluation Methods: User Survey

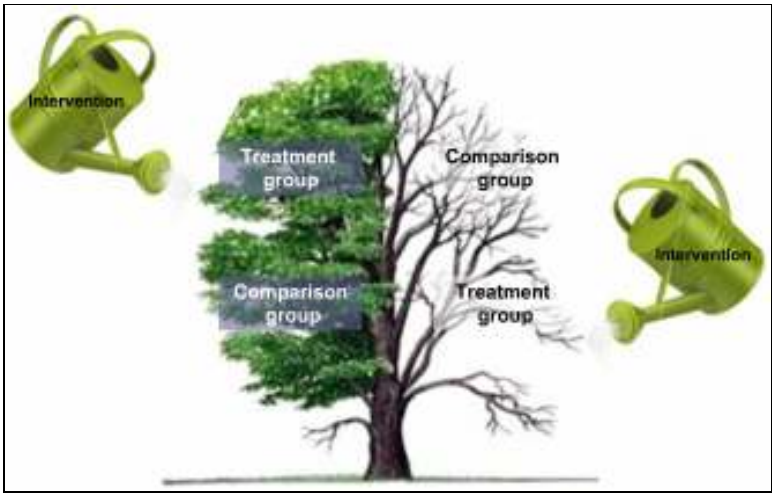
User surveys also are well adapted to the nature of innovation. They are able to provide a nuanced, quantified understanding of a policy and collect direct process experience as well as indicators. Finally they are able to test and generalize case study and other findings. On the other hand, they are frequently subject to positive bias, reflecting users' appreciation of receiving resources and optimism about impact.

Impact Evaluation: Field Experiments and Randomized Trials

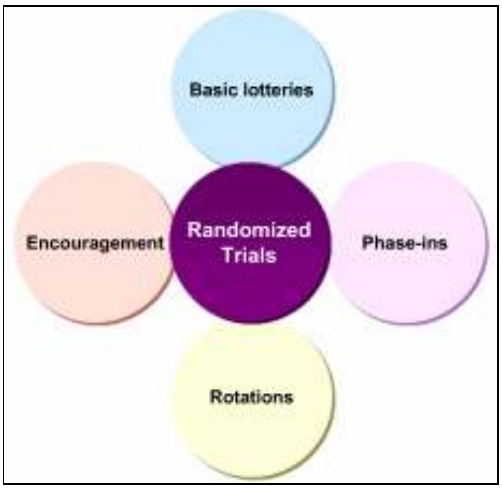
In recent years, there has been growing interest in the use of field experiments, especially randomized evaluations as an impact evaluation method. In natural sciences and some types of social research, they have become the "gold standard" in hard data production.



In particular, they go some way to remedying various selection biases, namely that systematic differences between comparison groups may be driving outcomes, and therefore provide a clearer indication of what would have happened in the absence of the intervention. They work by randomly allocating individuals to a treatment group of individuals which benefit from the intervention and a comparison group which do not, ensuring that any difference that arises can be plausible credited to the treatment rather than the characteristics of the individuals or other factors.



Randomized trials rely on number of methods -basic lotteries, phase-ins, rotations and encouragement- to safeguard the integrity of findings and can be designed to test the predictions of theory, the effectiveness of multiple policies and the conditions in which they might work.



Impact Evaluation: Field Experiment Examples

Field experiments have yielded some illuminating findings, including the importance of incentives for teachers, especially the importance of immediate rewards and the role of communication and social learning in the adoption of new technology. Because both program costs and data collection costs are relatively low in developing countries, it is possible to generalize beyond a few very well-crafted, major projects to a multiplicity of programs and contexts.

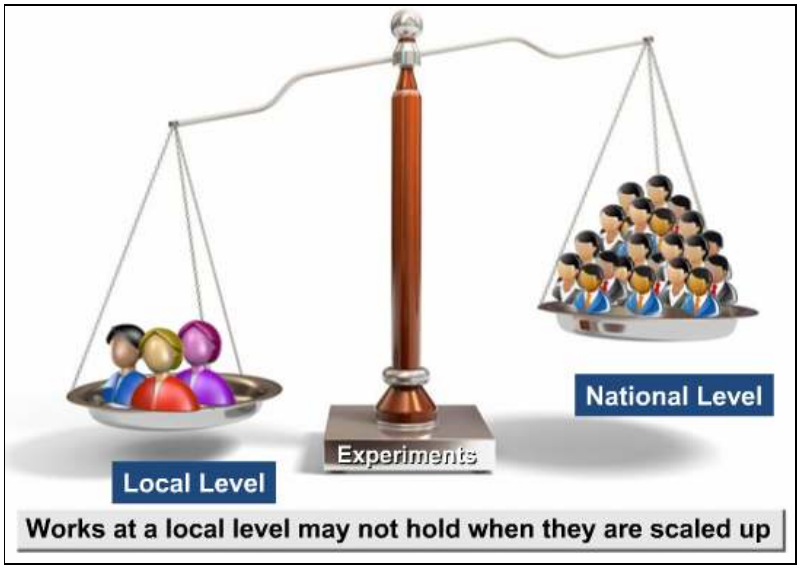
Limitations of Field Experiments and Randomized Trials

How far results can be truly generalized, however, is open to question. In order to draw credible conclusions about the general population, the sample size has to be sufficiently large. However, experiments tend to be quite specific in focus, reflecting the fact that they can only be carried out in a small number of locations or with specific partners.

Randomized trials are better at some tasks than others. While experiments might provide the correct answer about the behavioral response to an intervention, they are less equipped to explain why the intervention responded in a particular way. Any attempt to do so will necessarily invoke some theory or hypothesis in the background whose validity lies outside the findings of the experiment.

Field Experiments: Equilibrium Effects

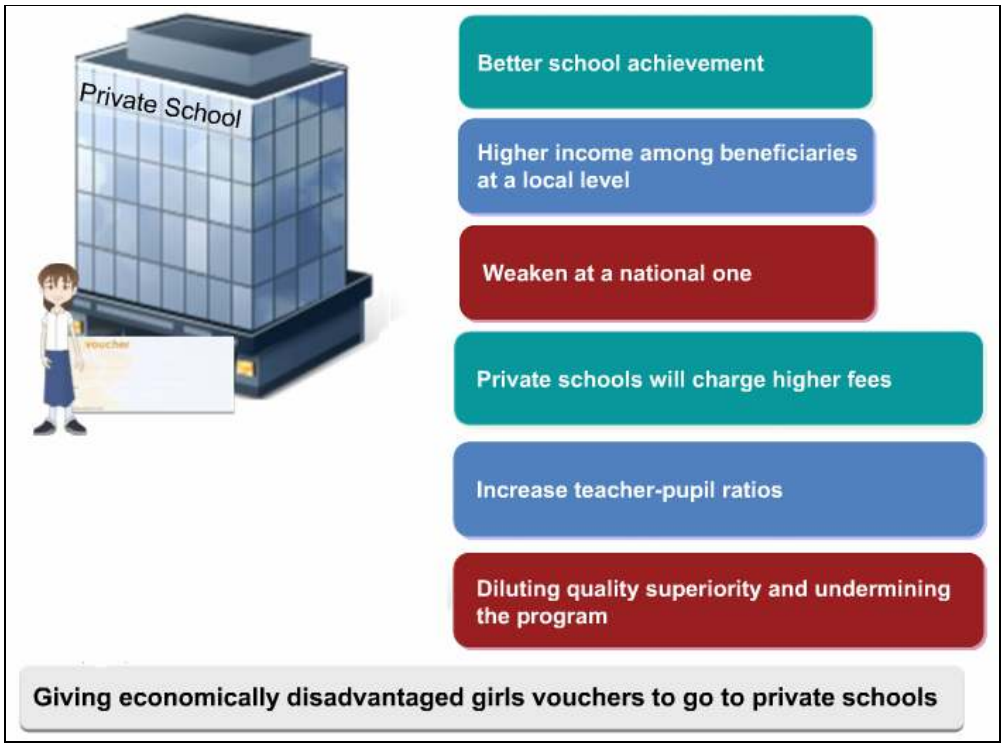
Groups may also be affected by their participation in the experiment or the unique characteristics of the implementing organization; similarly, so-called equilibrium effects may mean that what works at a local level may not hold when they are scaled up.



For instance, giving economically disadvantaged girls vouchers to go to private schools may contribute to better school achievement and higher income among beneficiaries at a local level; but this relationship may weaken at a national one: for instance, it is quite possible that private schools will charge higher fees, or if they are not allowed to do so, they may increase teacher-pupil ratios, thereby diluting their quality superiority and undermining the program. Similarly, scaling-up may have unintended consequences for the public school system; while the increased supply of educated girls owing to the pick-up of the scheme may serve to diminish the returns to education.

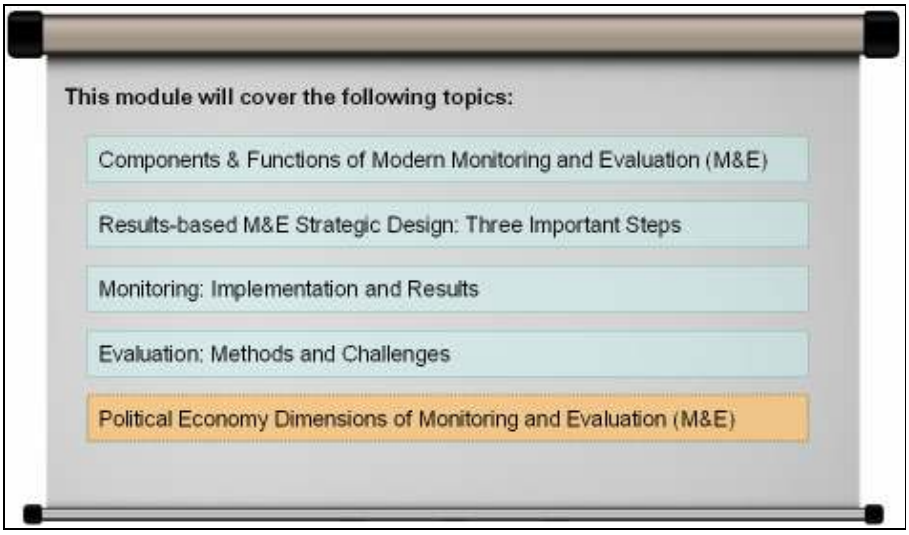
Finally, randomized experiments are most easy to run when looking at mean impact of programs, though these impacts are often the least interesting or insightful from a policy perspective.

Nonetheless, while randomized evaluations have their limitations, many of these either also apply to other quantitative methods or can be overcome with appropriate evaluation design.



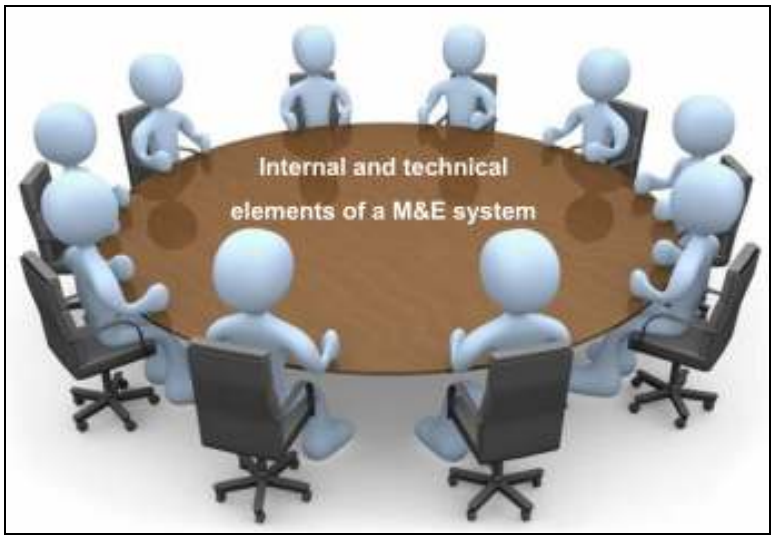
Module: Topics

And now we'll proceed to our next topic "Political economy dimensions of Monitoring and Evaluation (M&E) system"



Organizational Challenges to Building a M&E System

The discussion has so far focused on the internal and technical elements of a M&E system; but it must also live and breathe in a real politics setting and used by flesh-and- blood actors.



Indeed, a common problem is that M&E results often do not translate into policy design and implementation and are seldom used by policy makers at the strategic decision-making level. Sometimes this is a matter of expertise –bureaucrats and politicians are overwhelmed by information and lack the technical know-how to make sense of it. This can partly be addressed by deeper and more regular engagement between researchers and end-users at every stage of the project life cycle.

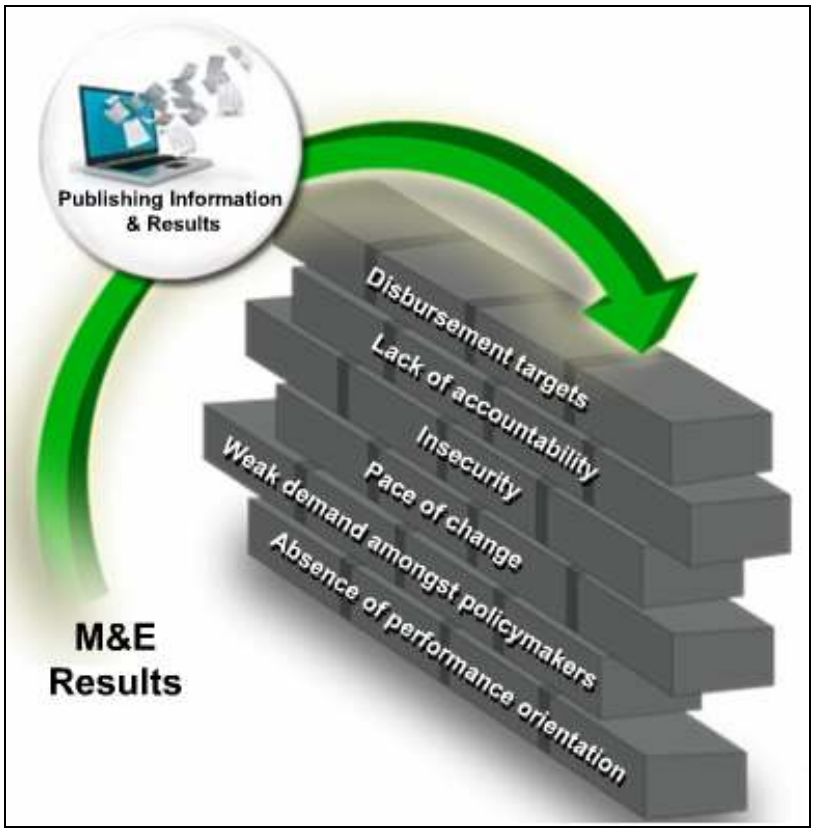
Cultural Challenges to Building a M&E System

But sometimes it is a matter of culture. It is important to acknowledge that technical data and evidence have to compete with other inputs as a basis for decision making. These include the experience and judgement of policy officials and politicians, values and ideology, available resources, habits and tradition, lobbyists, pressure groups and the media. These realities need to be thoroughly acknowledged –for instance, it is important for design and commissioning of M&E to coincide with the policy cycle; otherwise the principles of evidence-based policy may be granted less importance than they deserve. Evaluation approaches may also run into the "not invented here" syndrome – the tendency to reject ideas simply because they originate from outside the organization. As such, demonstrating the usefulness of evaluation principles to influential policymakers is of paramount importance.

Cultural Challenges to Building a M&E system: Organizational Learning

A proper culture of M&E requires a commitment to organizational learning. However, the pressure to spend and meet disbursement targets can lead to corners being cut during project planning and approval stages and important lessons being forgotten. Lack of accountability may also mean that actors have few incentives to learn from evaluations as they have generally moved on long by the time their errors have materialized. Finally, insecurity and the pace of change emanating from restructurings may lead to a loss of institutional memory.

More generally, weak demand amongst policymakers for M&E may reflect a weak evaluation culture, which in turns reflects the absence of performance orientation in the public sector. This implies that building a M&E system cannot be separated from wider efforts at modernizing the public sector.

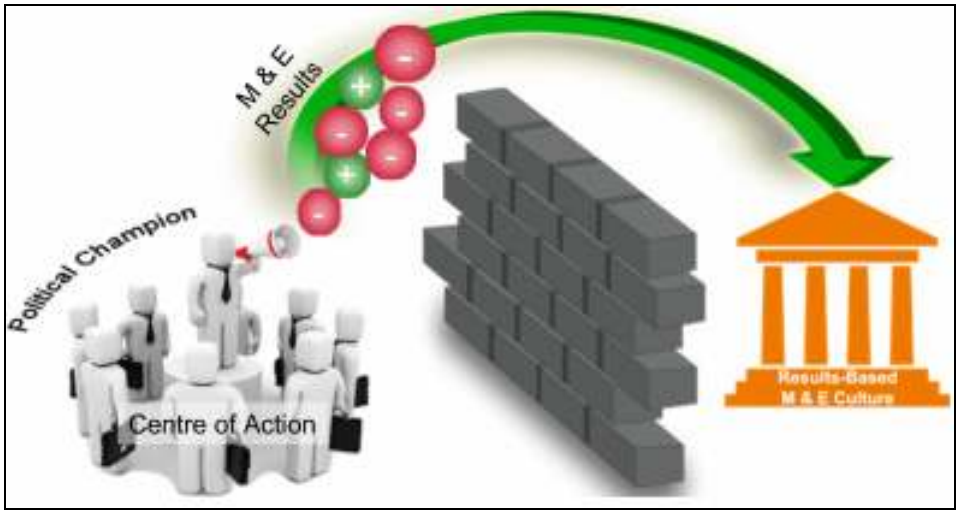


One strategy to ensure policymakers take M&E seriously is to publicize results. As a result policymakers are made aware that results will be scrutinized rather than be left to gather dust. Governments and organizations can use a wide array of strategies for sharing information with internal and external stakeholders. Periodic independent review by the national audit office, parliament, or a group of academics or civil society organizations to guarantee that the data generated by the system are accurate and reliable is the most obvious strategy.

Political Challenges to Building a M&E System

Finally, there may also be political-economy reasons why a results-based M&E system fail to take root in a given setting. Actors may have incentives to squash or manipulate information which is unflattering about their performance and potentially detrimental to their careers. Unless information is truthful, the M&E becomes a cynical, box-ticking exercise. Incentives have to be changed so that those who deliver bad news are rewarded rather than punished.

The success of M&E depends on a political champion to establish and defend such a system. This can be challenging for cross□cutting policies such as innovation which do not find a natural home in traditional, compartmentalized departments. It is not just enough that there exists a political champion for M&E; it must also be close to the centre of action so that its voice can be heard.



Conclusion

This module covered the following topics:

Components & Functions of modern approach to Monitoring and Evaluation (M&E) system

Results-based M&E Strategic Design: Three important steps

Monitoring of Implementation and Results

Evaluation: Methods and Challenges

Political economy dimensions of Monitoring and Evaluation (M&E) system