Evaluating Agricultural Innovation System Interventions

Andy Hall, LINK Ltd.
Kumuda Dorai, LINK Ltd.
Trish Kammili, National Institute for Agricultural Research, France (INRA)

SYNOPSIS

The evaluation of innovation system interventions is linked to the monitoring of interventions but is a separate function. Evaluation is usually performed on behalf of the investor by independent evaluators and not by those responsible for implementing the intervention. A well-designed evaluation is particularly important for innovation system interventions, as their process-driven nature means that the result chains are complex, dynamic, and not amenable to simple linear and anecdotal evaluation. Great care is needed in making judgments about the relationship between the effectiveness of the design and execution of the intervention and the observed impacts associated with it. Evaluation good practice is better equipped than ever to deal with this complexity. This note outlines key good practice principles relevant to evaluating innovation system interventions: a stronger learning orientation in evaluation, the use of counterfactuals, the use of mixed (quantitative and qualitative) methods, and incorporating the perspectives of different stakeholders. The example of Theory-Based Impact Evaluation (TBIE) illustrates how these principles work together. Experience with applying evaluation approaches and principles to agricultural innovation system interventions is currently in its infancy. A policy priority is to develop the capacity for this type of evaluation, which will help to improve the long-term effectiveness of innovation system interventions.

BACKGROUND AND CONTEXT

The effectiveness of innovation system interventions lies in their outcomes and impacts being investigated thoroughly and the resulting lessons applied to future investments. This process contributes to accountability as well as future investment performance. This note focuses on the evaluation of innovation system interventions.

Monitoring—the effective management of performance by implementers to achieve desired outcomes and report progress—is covered in TN 4.

Three critical features of AIS interventions influence how they are evaluated. First, they focus on strengthening capacity; second, they use a learning-based process; and third, they require a distinction to be made between impact, shared impacts, unexpected impacts, and unrelated impacts. Each of these issues is discussed next.

A focus on strengthening capacity

An innovation system intervention is an investment to improve how change and innovation take place. Although such an investment is made ultimately to achieve certain social, economic, or environmental impacts, the immediate outcome is improved ways of doing things—usually referred to as institutional change. For example, the National Agricultural Innovation Project in India established research, development, and private sector consortia around selected themes to introduce a new way of working that would achieve wide-scale impact in the future (see module 4, IAP 2). An innovation systems research project addressing fodder scarcity in West Africa and India (the Fodder Innovation Project, described in IAP 6) experimented with ways to induce institutional change that would enable innovation leading to social and economic impacts.

A learning-based intervention process

Improvements in how change and innovation take place will depend on the specific circumstances of each intervention. For this reason, ways of achieving those improvements are hard to design in advance. Innovation system interventions address this issue by adopting a learning-based approach,

in which the intervention continuously tests the assumptions on which it is based (its intervention logic). Over the course of the intervention, investors' expectations of what the intervention is going to achieve, and how it will do so, often need to evolve significantly. For example, an intervention in Sierra Leone (IAP 1) designed to help put agricultural research results into use began by trying to establish innovation platforms around commodity chains. It soon found that the main bottlenecks to research use and innovation were policy issues in the enabling environment for innovation.

Distinguishing impact, shared impacts, unexpected impacts, and unrelated impacts

By definition, innovation system interventions operate within an often complex web of activity. Much of this wider set of events is beyond the control and influence of the intervention, but the performance of the intervention itself is often greatly influenced by this context. While the immediate outcomes of an intervention—for example, the establishment of innovation platforms (see module 4, particularly TN 1)—can be evaluated easily, it is more difficult to attribute impacts to those platforms. The causal chain from intervention to impact is often complex, may take time to mature, and might be geographically removed from the point of intervention. As a result, tracing and attributing impact are difficult.

In India, for example, the Fodder Innovation Project (see IAP 6) introduced an institutional change, in which an NGO helped the National Department of Animal Husbandry use its resources to hold livestock health camps in villages. Can the resulting impact be attributed solely to the intervention that supported the NGO to explore institutional change? In Nigeria, the Fodder Innovation Project led to an unexpected outcome: An NGO and the local livestock research institute collaborated on animal disease surveillance, an activity that was outside the scope of the project's mandate to focus on fodder but was likely to lead to important future impacts. Finally, since complex impact chains can often mask underlying causal processes, there is the danger that impacts may be attributed to an intervention when in fact they arose from unrelated events.

INVESTMENT NEEDED

Viewed from the perspective of traditional approaches to assessing the impacts of agricultural research investments (see box 7.16), designing approaches for evaluating innovation system interventions seems daunting. Traditional approaches relied on (1) technology adoption studies, (2) investigation of economic surplus generated by research and computation of rates of return, and (3) economic studies of the contribution of research to impact. While these approaches are powerful (see Evenson, Waggoner, and Ruttan 1979; Pardey and Beintema 2001; Alston et al. 1995), a persistent critique is that they have weak diagnostic value. Their inability to elucidate underlying causal processes and account for institutional change are particular weaknesses (Hall et al. 2003; Horton and Mackay 2003; Watts et al. 2003).

In contrast, the wider development evaluation tradition is well equipped to deal with the investigation of causal links between dynamic theories of change and impacts of the sort likely to be encountered in innovation system interventions (see, for example, the guidance on impact evaluations in NONIE, Leeuw, and Vaessen 2009). The challenge is mainly to understand how current good practice in the wider development evaluation community can be more widely deployed in the evaluation of innovation system interventions. Of particular relevance is the recent focus on evaluations that link the assessment of outcomes and impacts (what were the end results) with learning (what processes and practices brought about those results) (Savedoff, Levine, and Birdsall 2006; White 2009a).

Principles for evaluation

Emerging from this trend is a set of principles that practitioners can draw upon in evaluating innovation system interventions. These are now general principles for all types of evaluations.

- ■ *Situational responsiveness influences the design of the evaluation.* Referred to as "situational responsiveness," the key principle involves matching the design to the needs, constraints, and opportunities of the particular situation rather than one particular method. The application of this principle rests on understanding the characteristics of an intervention or parts of it and determining where the underlying change processes are simple, complicated, or complex.

- ■ *Seek a stronger learning orientation in impact evaluation.* A number of new initiatives—notably the International Initiative for Impact Evaluation (3IE) but also others—have stressed the need to combine accountability and learning objectives. Learning is important for identifying what worked or did not work and why, and this information is valuable for designing future investments. While this perspective emphasizes the need for

Ex post impact assessment

- Although it addresses the accountability imperative for funders of interventions, it does not help in drawing the institutional lessons required to understand the innovation process. Does not cover noneconomic dimensions.
- Difficult to develop credible counterfactuals and establish definitive causalities.
- Focuses on intended positive results and frequently ignores unexpected and negative results.
- Not appropriate in more complex areas such as natural resource management, policy and biodiversity research, and training and capacity building.

Randomized control trials

- Weak in external validity (or generalizability) and in identifying the mechanisms responsible for differences observed in the experimental and control situations.
- Rarely appropriate in complex situations where outcomes arise as result of interactions of multiple factors that cannot be "controlled."
- Limited in their ability to deal with emerging and unanticipated outcomes.

Source: Kammili 2011.

Cost-benefit analysis

- Calls upon significant financial and human resources.
- Benefits are estimated in advance based on assumptions that may not always be correct.
- Not all costs and benefits can be quantified (social and environmental costs/benefits, for example); results obtained do not reflect all benefits.
- Items included in the analysis reflect the bias of whoever performs the analysis; coverage and quality thus vary greatly.
- Given its complexity, involves only economists and project designers and does not engage other primary stakeholders.

Economic surplus approach and rate of return studies

- Requires substantial resources for collecting, processing, and interpreting technical and economic data.
- Simplistic assumptions about lags, costs, and supply shifts have biased rates of return (usually upwards).
- Not suitable for ranking noncommodity research such as socioeconomic and interdisciplinary research.

rigor in measuring impact, it also explicitly acknowledges the need to test interventions' assumptions and theories of change (White 2009a). For details, see the example of Theory-Based Impact Evaluation in box 7.19.

- *Make assumptions explicit and revisit theories of change.* Adopting a greater learning orientation in evaluation means making assumptions about change processes explicit when planning interventions and revisiting and testing those assumptions at the time of evaluation. By exploring the validity of the assumptions, evaluators can learn critical lessons for designing future interventions (box 7.19).
- *Use counterfactuals in impact evaluation.* A stronger learning orientation has also been accompanied by demands for greater rigor in the quantitative methods used to measure impacts, as well as for methods that better establish what would have happened without a particular intervention—the counterfactual. The best way to achieve this goal is still a point of debate (see box 7.17 for details on the challenges involved). Box 7.18 provides an example of Propensity Score Matching, one approach to developing a counterfactual case.

- *Mixed methods.* Quantitative methods (to measure impacts) clearly are central to investigating impacts, but they must be complemented by a range of qualitative methods (to understand and learn from institutional and process changes) that can help to understand the context in which the intervention took place and the process to which the intervention contributed that brought about those impacts. Practitioners can draw upon a very wide array of qualitative methods, and the combination of methods selected depends considerably on the nature of the intervention and the precise demands of evaluation functions. (For a discussion of

The origins of counterfactuals can be traced to scientific experiments conducted with a control, the classic example being the randomized control trials used in drug testing. This approach poses ethical and operational problems that make it impractical for development interventions.

Another approach, "before" and "after" comparisons, struggles to account for changes that would have happened despite the intervention. Variants of this approach have been developed, however, which compare the effects of an intervention that starts in different places at different times. "With" and "with-out" comparisons appear more promising for capturing the counterfactual case, particularly when used in conjunction with statistical techniques that account for differences between the starting conditions in comparator sites.

The establishment of counterfactuals remains an evolving science for the evaluation community. Despite the challenges, mechanisms for establishing what would have happened in the absence of an intervention should be built into the overall design of interventions and the monitoring and evaluation arrangements that are put into place for them.

Source: Authors.

Propensity Score Matching is a tool for identifying a suitable group with which the recipients of an intervention (the treatment group) can be compared. Evaluators find a comparison group comprising individuals who did not, in fact, receive the intervention but who, given their observable characteristics, had the same probability of receiving it as individuals in the treatment group. The intervention's impact is the difference in outcomes between the treatment and comparison group.

Source: World Bank 2004.

mixed methods in Theory-Based Impact Evaluation, see box 7.19.)

■ *Incorporate different stakeholders' perspectives.* Greater participation is needed to capture the perspectives of the poor (and other stakeholders) on what were the actual results of the intervention (social, economic, and institutional). Similarly, wider participation is needed to more fully understand the process through which these results came about. Different stakeholders may have different interpretations of cause and effect. It is now good practice in evaluation to recognize these divergent perceptions. This principle extends to the need to draw on perspectives from a wider range of stakeholders in investigating the intervention's theory of change.

Evaluation approaches

Table 7.9 presents a comparative overview of evaluation approaches appropriate to the demands of innovation system interventions. In different ways and with different emphases, these methods use the principles discussed in this note. They should be viewed as a menu of approaches which practitioners can draw upon to ensure that evaluations achieve the correct balance between learning and accountability. There are a number of points that need to be highlighted about these methods. Only the first method mentioned, Theory-Based Impact Evaluation, explicitly makes provisions for measuring outcomes and investigating underlying process. In reality, Theory-Based Impact Evaluation is an evaluation framework rather than a specific method, and it relies on a suite of qualitative tools.

The other methods outlined in table 7.9 are specific tools for qualitative investigation of what happened and what processes lead to the outcomes observed. While these methods for learning lessons are powerful, practitioners must recognize that there is an appetite among investors—public and private—for the quantification of outcomes and impacts. The nascent approaches to evaluating innovation system interventions will need to satisfy this demand.

The least developed of the methods presented in table 7.9 is benchmarking of innovation capacity. An important area of methodological development is to find measures of

Box 7.19   Theory-Based Impact Evaluation

Theory-based Impact Evaluation (TBIE) is advocated for understanding why an intervention has or has not had an impact. The approach calls for examining and mapping the causal chain of an intervention—from inputs to outcomes and impact—to test the underlying assumptions and shed light on the "why" question. TBIE involves six steps:

1. **Map the causal chain.** The causal chain links inputs to outcomes and impacts. It is the intervention's theory of change, which explains how the intervention is expected to have its intended impact. Testing assumptions is central to a theory-based approach. One criticism of the causal chain approach is that it is static and does not account for an intervention's ability to adapt and evolve—an important consideration for innovation system interventions. This criticism is addressed in TBIE by suggesting that the intervention theory should reflect the new design, and the evaluation should document the learning process that resulted in the new design. In this way, the intervention theory remains dynamic. Another criticism is that it misses unintended consequences, which is addressed by identifying those consequences through a careful application of intervention theory, fed by thorough preliminary fieldwork, and using the resulting information to develop new theories of change.
2. **Understand the context.** Clearly, understanding the context is crucial to understanding its impact and in designing the evaluation, as context influences how the causal chain plays out.
3. **Anticipate heterogeneity.** Understanding the context also makes it possible to design the evaluation to anticipate possible variation in impacts (arising from how the intervention is designed, the characteristics of the beneficiaries, varied socioeconomic settings, and so on).
4. **Conduct a rigorous evaluation of impact using a credible counterfactual.** The appropriate counterfactual is most usually defined with reference to a control group, which has to be identified in a way that avoids selection bias, meaning the use of either experimental or quasi-experimental approaches. Panel data help to strengthen the design, so baselines are encouraged. Where they are not available, they might be recreated using existing data sets or recall. In addition to selection bias, important issues to consider in the design are the possibility of spillover effects (the control is affected by the intervention) and contagion or contamination (the control is affected by other interventions).
5. **Conduct a rigorous factual analysis.** The counterfactual analysis of impact needs to be supplemented by rigorous factual analysis of various kinds, given that many links in the causal chain are based on factual analysis. Targeting analysis is the most common form of factual analysis: Who benefits from the intervention? To the extent that there is a defined target group, then what is the extent of the targeting errors? Such errors can be quantified and their source identified. Factual analysis often highlights a crucial break in the causal chain and explains low impact.
6. **Use mixed methods.** A major step toward mixed methods is to increase the use of rigorous quantitative methods in qualitative studies (quantitative analysis informed by qualitative insight) or the use of qualitative data in quantitative studies. Without qualitative methods, the danger is that researchers will conduct impact studies with no exposure at all to the intervention.

Strengths

- The combination of counterfactual impact estimates with detailed exploration of causal links and theories of change has the potential to provide compelling lessons and evidence.
- Lessons gleaned from TBIE are valuable in trying to understand what works in development.
- Far more valuable lessons for policy are obtained through the insights on what doesn't work in interventions.

Weaknesses

- Not yet in widespread use, so experiences are limited.
- Construction of a robust counterfactual can be challenging.
- Data-intensive.
- Expensive, although good value for money if well executed.

Best use or application

- An evaluation tool best suited to investigating impacts and undertaking policy learning in large-scale interventions.

Source: White 2009b and authors.

innovation system performance at the macro level that rely on indicators of system behavior and functions rather than on input indicators such as research spending.

POTENTIAL BENEFITS

Effective evaluation is central to improving the performance of investments in current and future innovation system interventions. Well-designed evaluations of innovation system interventions are particularly important because their process-driven nature means that the result chains are complex, dynamic, and not amenable to simple, linear, anecdotal evaluation. Great care is needed in making judgments about the relationship between the effectiveness of an intervention's design and execution and the observed impacts associated with the intervention. Consequently the evaluation of innovation system interventions is likely to be expensive but will be an important investment in strengthening the long-term performance of these interventions.

POLICY ISSUES

Policy issues related to the evaluation of innovation system interventions are partly but not entirely similar to those for monitoring (see TN 4).

One similarity is the need to *build capacity in evaluating innovation system interventions.* Experience with these sorts of evaluations remains limited, because innovation system interventions are relatively recent. Nor are many investors familiar with the evaluation principles and approaches applicable to those interventions. Professional evaluators from the wider development community do have experience in applying these principles, but there is a need to develop a new cadre of evaluators with experience in applying them to AIS interventions.

A second policy priority is to *clearly demarcate responsibilities for monitoring and evaluation.* Investors need to make a very clear distinction between evaluation and monitoring and not be tempted to include the collection of impact evaluation data within the intervention. This point is particularly important for innovation system interventions, because the ultimate impacts may not emerge until some time after the intervention ends.

Third, *sufficient resources must be available to use a combination of evaluation methods.* The evaluation of innovation system interventions requires greater rigor in measuring impacts, investigating result chains, and testing the validity of theories of change. Inevitably, evaluations will take longer and cost more. Investors need to plan ahead and put sufficient money aside for evaluation when planning new innovation system interventions.

Finally, over and above the evaluation of individual interventions, investors will increasingly need to *benchmark innovation capacity developed in sectors and subsectors through innovation system interventions.* International investors may also wish to use benchmarking to make international comparisons. These comparisons are needed to track macro-level progress and to help target subsectors, sectors, and countries for investments to strengthen innovation capacity or for other investments that require certain levels of innovation capacity as a precondition. A number of methodological challenges in measuring and comparing context-specific and systemic capacities of this sort remain to be addressed, however.

LESSONS AND RECOMMENDATIONS FOR PRACTITIONERS

The main lessons and recommendations for practitioners mirror the principles of good practice outlined earlier. A description of Theory-Based Impact Evaluation (box 7.19) illustrates how these good practice principles and the methods associated with them work together to ensure that the evaluation of innovation system interventions contains accountability and learning dimensions.

The evaluation of innovation system interventions should give particular attention to the following:

- *Timing of evaluations.* While many innovation system interventions will have quick wins, most impacts will emerge only much later as institutional change kicks in, creating new capacities for innovation. While process and institutional change evaluation can take place shortly after the completion of an intervention, impact evaluation can take place only after sufficient time has passed, often 3–5 years later.
- *Looking widely for impact.* The evolving nature of interventions means that outcomes and impacts can be unpredictable, both in the types of impact observed and their geographical and social location. Evaluation needs to be sensitive to this unpredictability.
- *Investigating unexpected outcomes.* Since innovation system interventions operate in dynamic environments, changes and unexpected outcomes can occur that have significance for impacts or can provide opportunities for new investments. These unexpected outcomes need to be recorded and investigated for any lessons they provide.

Table 7.9   Evaluation Approaches Relevant to Innovation System Interventions

Method	Purpose/use	Strengths	Weaknesses	Major references
Theory-based Impact Evaluation	– Helps in understanding why a program has or has not had impact. – Best suited to investigating impacts and undertaking policy learning in large-scale investments.	– A combination of counterfactual-based impact estimates with a detailed exploration of causal links and theories of change has the potential to provide compelling lessons and evidence. – Lessons gleaned from such an exercise are valuable in trying to understand what works in development. – Offers far more valuable lessons for policy through its insights on what doesn't work in development.	– Not yet in widespread use, so experience limited. – Construction of a robust counterfactual can be challenging. – Data-intensive. – Expensive, though good value for money if well executed.	White (2009b)
Innovation and Institutional Histories	– Understand past innovation processes and identify institutional factors that foster or hinder innovation. – Forge a shared vision of the future among stakeholders.	– Fosters discussion among stakeholders and leads to reflection and learning. – Helps build a shared vision of the future.	– Written accounts are based on recollections and could be biased. – Need skilled facilitators to help collect and analyze the information. – Usually takes a very long time for significant lessons to emerge.	Douthwaite and Ashby (2005); Shambu Prasad, Hall, and Thummuru (2006)
Participatory Impact Pathway Analysis	– Guide project management, especially in complex situations where innovation is seen as emerging from a network.	– Involves intervention staff and key stakeholders in constructing impact pathways. – Includes both (1) a causal chain of activities, outputs, and outcomes that sheds light on how an intervention achieves its goals and (2) network maps that show evolving relationships between participants. – Underlines the fact that innovations emerge from a network and not a linear "pipeline." – Promotes reflection, self-evaluation, and learning. – Provides a framework for carrying out action-research.	– A relatively new approach that needs to be further implemented to gauge its weaknesses.	Douthwaite et al. (2003)
Causal Process Tracing	– Well suited for complex, long-term interventions with systems learning goals.	– Places data and theory in close proximity. One quickly sees what works and what does not in an intervention's lifetime.	– Regarded as not very strong for wider generalization but more suited to narrow specification of the reach of causal propositions. – Takes a great deal of time. – Not conducive to parsimonious theory and leads to partial, middle-range theory. It is easy to miss causal complexity. – Easy to lose sight of the broader context.	George and Bennett (2005); CoS-SIS (2009); Walters and Vayda (2009)

Most Significant Change	– Make sense of an intervention's impact and foster learning. – With the help of all primary stakeholders, identify the most significant changes that have occurred as a result of an intervention.	– Enhances capacities of stakeholders (at the organizational and individual levels) and fosters learning. – Helps identify unexpected changes or outcomes. – Large amounts of information are processed, from which negative and positive changes are deduced. – More accessible than traditional M&E techniques; no specific skills required to participate.	– A subjective expression of the values and concerns of the stakeholders designated to select the stories.	ECDPM (2006); IFAD (2002); Davies and Dart (2005)
Stories and Narratives	– Shed light on the changes that have occurred at the individual, organizational, or institutional level. – Uncover intangible factors (qualities, values, culture, and so forth) that determine the organization's character.	– Exchanging stories builds trust between participants; in some cases may lead to an environment conducive to learning and stimulate change. – An effective way to deal with passions and emotions of individuals involved. – Has the potential to stimulate change if told correctly. – Individual focus can translate into organizational development. – Not hierarchical.	– Biased; depends on the perspective of the person telling the story.	Asif (2005)
Benchmarking Innovation Capacity	– More than just evaluating individual interventions. Can help investors benchmark the performance of sectors and subsectors in terms of capacity built for innovation.	– Can be useful in tracking macro-level progress and to help target subsectors, sectors, and countries for investment in capacity strengthening or for investments that require certain levels of capacity as a precondition.	– Methodologies still being developed.	CPR, CRISP, and LINK (2008); Kraemer-Mbula (2012 forthcoming); Spielman and Birner (2008)

Source: Authors.

■ ***Acting on evaluation lessons.*** Investors must be willing to act upon the lessons that emerge from evaluations of their innovation system interventions. This can be easier said than done when lessons point investors away from traditional approaches and the interests of stakeholders associated with them. One approach that investors increasingly use to bolster their confidence in emerging lessons is to use systematic reviews of similar projects to test key assumptions about suggested ways forward.

The principles outlined in this module and the evaluation approaches summarized in table 7.9 are well known to professional evaluators in the wider development evaluation community. What is important is that investors commission evaluations that embody these principles and select evaluators with experience in applying them. In the short term, they probably must look beyond the pool of evaluators who have undertaken traditional assessments of the impact of agricultural research investments.