Policy BRIEFING

ISSUE 78 • OCTOBER 2014

Agricultural Impact Evaluation is Failing to Measure Up

New technology that enables sustainable and profitable production of food is critical for both food and nutrition security and economic development. Yet, recent research suggests assessments of the productivity gains farmers realise from new technology are routinely flawed methodologically and hence unreliable as a basis for decision making. As a result, opportunities to support this key aspect of agricultural performance and more equitable benefits from it have been missed. This briefing explains why and highlights measures to stimulate demand for methodological quality in evaluations and reinforce their contribution to strengthening systems of innovation.

Assessing the gains from new technology

Whether development objectives are framed in terms of poverty reduction, environmental protection, adaptation to climate change or modernisation, technological change is at the heart of most agricultural policy and programmes. The distribution of increased productivity is a key developmental concern and a nagging question is why the benefits of new agricultural technology often appear to by-pass poorer farmers – even when they are the 'target' group.

In 2013, IDS and Africa Rice researchers concluded a **systematic review** of technologies for food crop production in low and lower middle income countries (LLMIC) and the conditions and circumstances in which farmers achieve productivity gains when adopting them. It was also concerned with other impacts, positive and negative, that may accrue, for example with respect to health, food security or environmental services. Their key objective was to provide policy makers and practitioners with a more realistic understanding of the outcomes they can expect from technological change and the opportunities to shape the innovation environment so as to favour more productive agriculture.

Researchers screened a total of 20,299 papers that emerged from their search of the literature, among which they found 214 that were broadly relevant to the review question. A close reading of these papers found that 209 failed against two of the quality criteria they had established: clearly defining technology 'adoption' and employing a clear and credible method for assessing change in productivity. Only five papers, scattered across ten technology groups, passed all the criteria.

Five papers is too narrow a body of studies on which to base a meaningful synthesis of evidence. The researchers were therefore unable to complete the systematic review and conclude anything about the conditions and circumstances influencing the productivity gains farmers achieve from new technology – the core question that drove the research – because in the large majority of studies it wasn't clear whether and what benefits were realised or as a result of what.

The key finding lies not in anything one could take from the few studies that passed the screening but in the fact that so many did not. The study concluded that the assessment of productivity change as a result of technology adoption in much of the published literature is methodologically flawed. As a result, any attempt to derive clear, evidence-based guidance on the conditions and circumstances

"If there is demand for quality evidence to inform policy and programmes, why is it so ineffectively expressed in the evaluation process?"

under which farmers achieve productivity gains when they adopt technology is undermined. Opportunities to support more effective policy and programmes in relation to this key aspect of agricultural performance have been lost and, it would appear, a good deal of research time and money wasted.

Gaps in evaluation governance

The two criteria that most papers failed to meet were neither difficult nor only recently recognised. It has long been understood that adoption is more than a simple yes/no decision, that three dimensions of use affect the productivity and other outcomes that can be expected from a technology: how long farmers have known and used it, on what area of their fields they use it and how much of it or what parts of it they use. A guide to the design of technology adoption studies was published and widely disseminated 20 years ago. Similarly, feasible and accessible methods to assess change in productivity are described in text books and taught in many undergraduate and professional curricula. The criteria the study applied were by no means unrealistic.

The box opposite describes one of the studies that failed to meet these quality criteria – involving the impact of new crop cultivars in eastern Africa – and outlines why it was judged to have failed and what straightforward and inexpensive modifications to their methods would have allowed it to pass. Other solutions are also possible.

The key question this review raises, is a systemic one: if obtaining quality evidence is feasible, why is it so often not supplied by evaluations in this area? Or, looked at from the other side of the market: if there is demand for quality evidence to inform policy and programmes, why is it so ineffectively expressed in the evaluation process?

If that demand is to become more effective, the incentives that militate against it being supplied in the current institutional environment must first be recognised. For the organisations engaged in research on and development and diffusion of new technology, and for the individuals directly responsible, evaluation often serves to confirm 'success' and secure profile in a competitive environment. Accountability is predominantly upwards, to their funders. In this context, evaluations often take on a perfunctory or symbolic role, demonstrating compliance with institutional directives and adherence to academic standards. However, this context also insulates evaluation from accountability to those actors much closer to where technologies are actually used, adapted and replaced who arguably have the greatest stake in evaluation's quality and relevance.

The reasons why quality in impact evaluation often appears to be so little valued are closely related to the reasons why impact evaluations have not made more of a difference in agriculture. This was the subject of a conference more than a decade ago where several contributors saw the failure to situate impact evaluation within innovation systems as the core of the problem. A key feature of innovation systems, which contrasts with the dominant, linear perspectives on technological change, is that all actors are understood to initiate innovation and that actors relate to one another in more than a passive or hierarchically determined fashion: interactive learning is central to the well functioning of a system of innovation.

Evaluation can be the cornerstone of this interactive learning. To fulfil this role, the focus must broaden from the individual organisation to include the network or coalition within which it works, which is jointly innovating towards common objectives. In this context, evaluation quality, in an inclusive sense, is more likely to be assured and its findings stand a greater chance of being directly used. When other actors within the system of innovation are involved in the planning and conduct of the evaluation, they can influence the questions it addresses, enhancing its relevance. Negative or unintended outcomes are more difficult to hide or avoid, diminishing the reporting bias that dogs evaluation when demonstrating short term success is a key motivation. Actors who know the terrain may be able to offer alternative or complementary explanations for the findings: discussion around these can make the evaluation more transparent and legitimate. In the example described in this briefing, local actors might have pointed out that the comparison on which the assessment of productivity change was based involved crops growing in very different conditions. Such feedback can support methodological rigour appropriate to the situation. Where relationships between actors are ongoing, the incentives and opportunities to improve evaluation quality are clearer.

"The reasons why quality in impact evaluation often appears to be so little valued are closely related to the reasons why impact evaluations have not made more of a difference in agriculture."

Choosing and using new crop varieties

Common bean (*Phaseolus vulgaris*) is a key nutritional resource and an important source of income in much of East Africa, particularly for poorer farmers. Research has aimed at increasing the productivity and profitability of the crop through the development of sustainable technologies, notably diverse varieties tolerant of a range of stresses.

In three neighbouring villages in a major bean growing district of Uganda, researchers made small packets of seed of two new varieties available to farmers through local market vendors and women's groups. They assessed how farmers valued and used the varieties and what outcomes they achieved through a baseline and a follow-up survey three years after the introduction. Other investigations provided insight into the food security benefits of the new varieties and the extent of their spread in other villages in the district.

Defining adoption

The follow-up survey found that most users of the new varieties had been growing them for 2-3 years and had taken them up in the first two years after they became available. Across wealth classes, they planted them on the majority of the land they devoted to beans: in aggregate, close to three quarters of the villages' bean area.

The study adequately characterised adoption in temporal and spatial terms. The proportion aspect of adoption wasn't relevant here: as generally with new varieties, there were no parts of the technology to choose from.

Assessing productivity change

Researchers calculated the productivity advantage of the new varieties as the difference between their yields and those of one or both of two formerly common varieties that were grown on the same farms, as reported in the follow-up survey. These advantages were 35 per cent and 38 per cent for one of the new varieties and 69 per cent and 79 per cent for the other.

This study was excluded for not adequately assessing productivity change. The comparison of yields in one season doesn't provide a credible estimate of the increase in yield that farmers achieved by taking up the new varieties. Farmers reported that they had substantially reduced the area on which they planted the older varieties: they may well also have relegated them to less favoured parts of their fields, which would exaggerate the new varieties' yield advantage. Qualitative evidence indicates that increased productivity played a part in farmers' decisions: in interviews, almost all said they appreciated the new varieties for their high yield. However, other characteristics were almost as frequently cited such as drought tolerance, marketability and cooking time. Experience elsewhere with beans and other crops shows that farmers typically consider multiple criteria in their choice of varieties.

How might it have been done differently?

The researchers carried out a baseline survey but didn't measure or inquire about yields. Had they assessed yields in the same fields as in the follow-up survey, they could have used a difference-in-difference analysis to compare the change in yield for farmers who had or hadn't taken up the new varieties. The analysis could have controlled for wealth and gender of the household head. Poor and female-headed households had been reported less likely to adopt the new varieties: it would have been revealing to see if these factors – 'conditions' in the terminology of the systematic review – also influenced productivity gain.

The research amassed detailed information about the farmers who grew these varieties, the networks and markets through which seed and information spread, and the outcomes in addition to yield that farmers achieved and valued. It is unfortunate that it didn't provide a clearer picture of whether and by how much farmers actually increased their production and how that gain was distributed.

"When other actors within the system of innovation are involved in the planning and conduct of the evaluation, they can influence the questions it addresses, enhancing its relevance."

Recommendations

The widespread yet avoidable flaws in impact evaluation that the systematic review revealed point to a pervasive lack of demand for quality. The discussion in this briefing suggests that this demand is more likely to be sustained when evaluation is situated within the networks and coalitions in which technologies are developed, adapted, used, and their outputs commercialised. Situating evaluation there will, at the same time, contribute to the interactive learning that underpins a well functioning system of innovation which is essential for technological change and adaptation to changing conditions.

A range of actors can contribute to creating conditions in which demand for quality evaluation is more effective:

- **Stakeholders in agricultural innovation systems at national and international levels should support institutional cultures less fixated on short-term 'success', in which evaluation is understood as an essential tool to assess organisations' work, learning from and improving on it.** One consequence would be to reduce the current disincentives for quality in impact evaluation.
- **Funding agencies can encourage the broadening of evaluation focus from organisation to networks and coalitions by the terms they agree for evaluations with executing agencies.** They should ensure actors in these networks and coalitions are involved at all stages of evaluation. These actors should also press for their inclusion in these evaluations since they relate to their ability to work effectively.
- **The outcomes evaluations assess should not be limited to those such as productivity increase that have commonly been a dominant concern of funding agencies but include as well those that other stakeholders value.** Among these are likely to be indicators of the health of linkages between actors in innovation networks and coalitions. This would serve to reduce perceived contradictions between and discordant pressures for upward accountability and learning.
- **Demand for quality can be supported by more effective review processes alert to the kind of methodological flaws this study has revealed:** in the internal review of proposals and reports and in the peer review of papers submitted for publication. This would provide users greater assurance that evaluation results are a reliable basis for decisions.
- Actions on the supply side can respond to demand for quality evaluation. **Organisations should, as far as they are able, enlarge their internal capacity for evaluation.** This would help in making evaluation a more accessible tool for assessment and improvement. Greater use should be made of mixed quantitative and qualitative methods and the social science skills to adapt them should be strengthened.

Institute of Development Studies

UKaid from the British people

IDS Policy Briefings are published by the Institute of Development Studies and aim to provide high quality analysis and practical recommendations for policymakers on important development issues.
To subscribe: **www.ids.ac.uk/idspolicybriefings**

Institute of Development Studies, Brighton BN1 9RE UK
T +44 (0) 1273 606261 **F** + 44 (0) 1273 621202 **E** ids@ids.ac.uk **W** **www.ids.ac.uk**
twitter.com/IDS_UK #idspolicy facebook.com/idsuk

Further reading

Guijt, I. (2010) 'Exploding the Myth of Incompatibility between Accountability and Learning', *Capacity Development in Practice*, J. Ubels, N-A Acquaye-Baddoo, and A. Fowler (eds), London, Washington: Earthscan

Hall, A., V. Rasheed Sulaiman, N. Clark, and B. Yoganand (2003) 'From measuring impact to learning institutional lessons: an innovation systems perspective on improving the management of international agricultural research', *Agricultural Systems* 78(2):213-41

Horton, D., and R. Mackay (2003) 'Using evaluation to enhance institutional learning and change: recent experiences with agricultural research and development', *Agricultural Systems* 78(2):127-42

Stern, E., N. Stame, J. Mayne, K. Forss, R. Davies, and B. Befani (2012) 'Broadening the range of designs and methods for impact evaluations', *Working Paper* 38, London: UK Department for International Development

Loevinsohn, M., Sumberg, J., Diagne, A., Whitfield, S. (2013) 'Under what circumstances and conditions does adoption of technology result in increased agricultural productivity?' A Systematic Review Prepared for the Department for International Development, Brighton: Institute of Development Studies

Credits

This *IDS Policy Briefing* was written by **Michael Loevinsohn** and is based on a systematic review conducted by Michael Loevinsohn, Jim Sumberg, Aliou Diagne and Stephen Whitfield. The editor was Carol Smithyes.

ISSN 1479-974X