

TOOLS TO ASSESS VALUE FOR MONEY IN AGRICULTURE

INTRODUCTION

In sub-Saharan Africa, 80 per cent of the population rely on agriculture for their livelihoods, and they do so with low levels of agricultural productivity. Agriculture is critical in food security, poverty reduction and overall economic growth. For many years, agricultural productivity in sub-Saharan Africa has lagged behind other regions such as South Asia, East Asia and the Pacific. In order to raise productivity levels in agriculture, several strategies have been employed by African governments that have led to better performing economies. An example of this is the increased cotton production in Burkina Faso, which was an important factor in doubling real GDP per capita between 1995 and 2006 (Kaminski 2011).

Governments have to use their scarce resources for various equally important priorities. The agriculture sector is no different: a limited budget needs to be prioritised over a number of commodities, inputs and sub-programmes. With a better understanding of the relative returns of various investments, agriculture policies can target resources towards activities that generate good value for money. It is crucial, therefore, that governments allocate their budgets effectively and efficiently in the highest-return agricultural projects that, in turn, will help stimulate growth and decrease poverty.

This paper examines the importance of impact evaluation tools and techniques. The paper also discusses how these tools can be applied and institutionalised in the public sector.

WHY EVALUATE AGRICULTURE PROJECTS?

The New Rice for Africa (NERICA) programme is a good example of why and how impact evaluations matter for evidence-based policy-making. NERICA rice¹ was first introduced in 1996 in Côte d'Ivoire and has since spread rapidly

across many sub-Saharan African countries. By 2006, 200 000 acres of land were devoted to NERICA varieties in the whole of sub-Saharan Africa. Various impact evaluations have been carried out since the introduction of the programme, and the effect of the adoption of these new varieties of rice on poverty, household wealth, productivity and economic growth has been documented. Kijima, Sserunkuuma and Otsuka (2006) record the impact of NERICA adoption on average yield in Uganda, and find that average yield of NERICA was twice as much as the average rice yield in Africa. Kijima, Otsuka and Sserunkuuma (2008) show that NERICA increased per capita income by USD20 and decreased the incidence of poverty by 5 percentage points. The success of this programme has been due partly to the extensive effort put into the evaluation of its impact. The success of the NERICA Rice Dissemination Project in Mali, for instance, led to the adoption of this new technology in Benin. Along the way, the programme was redesigned and re-evaluated in order to achieve the best possible outcome.

There is a growing global trend towards evidence-based policy-making. Policy-makers are interested not only in controlling the inputs and activities of programmes, but also in determining if programmes have achieved their intended goals. Evaluation (along with monitoring) is at the heart of evidence-based policy-making. Monitoring is a continuous process whereby inputs, activities and outputs are tracked on a daily basis. Evaluation, on the other hand, is a periodic review of a planned, ongoing or finished programme.

Understanding the impact and benefit of a programme helps policy-makers to make better allocative decisions. Policy-makers then can take steps to improve the effectiveness, efficiency and quality of a programme at various stages of the implementation process. The estimated impact of the programme will be used to calculate its expected benefit, and will provide critical inputs for a cost-benefit analysis. Advancing such practices across the public sector will lead to an improvement in policy-making and service delivery.

¹ Hybrid rice varieties developed by the Africa Rice Center (AfricaRice) to improve the yield of African rice.

Why evaluate the impact of development projects?

Many governments, institutions and project managers are reluctant to carry out impact evaluations because they are deemed to be expensive, time consuming and technically complex, and because the findings can be politically sensitive, particularly if they are negative. Many evaluations have also been criticised because the results come too late, do not answer the right questions, or are not conducted with sufficient analytical rigour. A further frequent constraint is the limited availability and quality of data. Yet, with proper and early planning, the support of policy-makers and a relatively small investment (compared with overall project cost), a rigorous evaluation can be very powerful in assessing the appropriateness and effectiveness of programmes. Evaluating impact is particularly critical in developing countries, where resources are scarce and every dollar spent should aim to maximise its impact on poverty reduction. If programmes are poorly designed, do not reach their intended beneficiaries or are wasteful, with the right information they can be redesigned, improved or eliminated if deemed necessary. The knowledge gained from impact evaluation studies will also provide critical input to the appropriate design of future programmes and projects.

Source: Baker (2000)

OVERVIEW OF IMPACT EVALUATION METHODS

Impact evaluation techniques are divided into groups, depending on the type of data being used, the timing of the evaluation and how participation is assigned. Studies in the literature are either qualitative or quantitative, either *ex post* or *ex ante* and either experimental or non-experimental.

Qualitative impact evaluations delve into mechanisms of impact and are generally more helpful in understanding the underlying processes through which the programme succeeds or fails to achieve the intended outcome. Quantitative studies require less contextual methods and apply rigorous statistical and econometric analysis; such data and analysis, in turn, provide researchers with conclusions that are deemed more reliable. There is, however, an increasing trend in the literature towards the mixing of qualitative and quantitative methods.

Ex post evaluations (e.g. randomised controlled trials, differences in differences, matching, regression discontinuity design and instrumental variables estimation) look retrospectively at the impact of programmes that have already been implemented. *Ex ante* evaluations (e.g. structural economic models and simulation techniques), on the other hand, are conducted either before the programme starts or before it changes in some way; in other words, such evaluations aim to provide forecasts of the potential impact of

a programme that is yet to be implemented or of alternative applications of an existing programme.

Several issues determine which technique should be deployed. Depending on the question that is posed by the evaluator, the timing of the evaluation, actual design of the programme and data availability would determine the technique that is selected. Given the dominance of *ex post* evaluation techniques in impact evaluations, due to their technical rigour, three of the most commonly used tools are presented below with illustrative examples.

Randomised control trials (RCTs)

RCTs are experiments that are designed specifically to measure the impact of a certain programme. An RCT requires a random assignment of the observation units (e.g. farmers) into treatment or control groups. Treatment groups (also called participants) are those who receive the intervention (e.g. a subsidy). Control groups are similar in characteristics to the treatment group, except that they do not receive the intervention. Since the allocation of treatment is random, there is no reason why the groups should be different from one another in observable and unobservable characteristics. In this sense, RCTs are said to meet the golden standard of evaluation: they produce a good estimate of the counterfactual² and remove any bias due to a selection problem.³ If the RCT is successful in building suitable control and treatment groups, the observable characteristics should not be statistically significantly different between the control and the treatment groups. If so, any difference in the outcome variable that emerges between the control and the treatment groups after the implementation of the programme will be an unbiased estimate of the programme's impact.

An example of this tool can be found in Duflo, Kremer and Robinson's (2008) examination of whether fertiliser usage increases profits and, if so, under what conditions the benefits of fertiliser usage are maximised. A series of field trials was conducted in a relatively poor district in Western Kenya. Farmers were randomly selected and allocated plots. One plot was randomly selected to receive a specific fertiliser, whereas another one was given the full package recommended by the Ministry of Agriculture. There was a third plot, which served as the control group. During the implementation of the intervention, farmers were assisted in different ways as to how they should be using the fertiliser and the new hybrid seeds. Duflo et al. (2008) conclude that although fertiliser can be very profitable, there is substantial scope for learning, and small (and inexperienced) farmers should be supported in different ways for such profits to be realised. This study is a good example of an RCT where there is more than one treatment group, and the intervention is

2 The state of the world in the absence of the intervention. For most impact evaluations, the counterfactual is the value of the outcome for the treatment group in the absence of the intervention. However, studies should also pay attention to unintended outcomes, including effects on non-beneficiaries.

3 Potential biases introduced into a study by the selection of different types of people into treatment and comparison groups. As a result, the outcome differences may potentially be explained as a result of pre-existing differences between the groups, rather than the treatment itself.

not a 'one-shot' type intervention but a series of interventions over a longer period of time. This allowed the researchers to answer many questions at once and also to understand the reason why some farmers benefit more/less from fertiliser usage. The bottom line was that fertiliser usage should be complemented with technical support. It was the way in which the experiment was designed that allowed the researchers to reach this conclusion.

Difference in differences (DID)

DID is a popular evaluation technique that is applied to observational data. It is a quantitative, non-experimental evaluation method, which combines the before-after and participant-nonparticipant comparison techniques in a novel way. The idea is as follows: if we have data for the control and the treatment groups for both the pre- and post-implementation periods, we can do double differencing to get an unbiased estimate of a programme's impact. First, we get the difference between the control and the treatment groups before the implementation of the programme. Second, we get the difference between the same control and treatment groups after the implementation of the programme. Finally, the difference between the two differences gives us an estimate of the impact of the programme. The underlying assumption here is that the difference between the control and the treatment groups would have remained the same had the programme not been implemented. Any change in the difference between the control and the treatment groups from the pre- to the post-intervention period is due to the implementation of the programme being evaluated.

Deininger, Ali and Alemu (2009) offer an example of this technique. Their study examined the effects of a low-cost land-registration programme in Ethiopia on land-related investment. They used data from four rounds of a panel survey of rural households conducted in 1999, 2001, 2004 and 2007. The fact that the programme was introduced at different times in different districts made it possible for the researchers to adopt a DID approach. The main outcome variable was land-related investment. They also looked at the impact of the programme on participation in the land rental market, and expectations of conflict in the next five years. They conclude that certification had a positive economic impact, and improved tenure security and the supply of land to the rental market.

Instrumental variable estimation (IVE)

The main problem faced while conducting an evaluation is to find a way of solving the selection bias problem. That is, if some unit characteristics that determine participation are also correlated with the outcome variable, a simple comparison of the control and the treatment groups will produce biased estimates of programme impact. IVE solves this identification issue by introducing a third variable, which is called the instrumental variable. For IVE to produce unbiased coefficients, the selected instrument has to satisfy two properties. First, it needs to be correlated with the variables measuring

participation. Second, the only correlation between the instrument and the outcome variable should be due to and through the variable measuring participation. If the instrumental variable meets these two properties, it will be a valid instrument. The strength of this method is due to the fact that with a valid instrument, IVE can control and correct for unobserved as well as the observed differences between the control and the treatment groups.

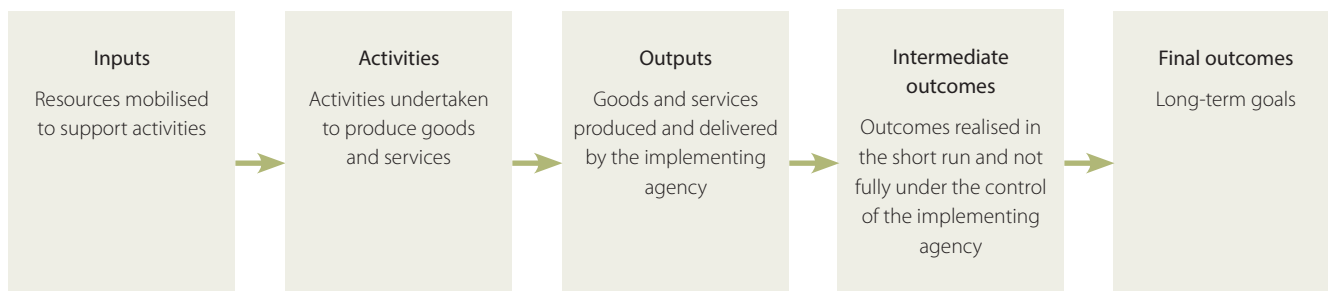
Ricker-Gilbert and Jayne (2010) studied the contemporaneous and dynamic impacts of fertiliser subsidies on household well-being. They realised that non-random distribution of targeted distribution subsidies might potentially lead to selection bias in a simple regression of measures of well-being on fertiliser usage. In order to overcome this, the authors applied IVE. Their instrument was a variable measuring whether or not a member of Parliament resided in the community. In order to convince the reader that this is a valid instrument, they assert the following: firstly, that variable is a measure of socio-political capital that could influence how much subsidised fertiliser a household receives; and, secondly, one does not expect to see a direct effect of a member of Parliament residing in the community on the well-being of the household through channels other than access to fertiliser. This second assertion is questionable and harder to test, which may cast doubt on the results of this paper. One objection could be that if a member of Parliament resides in a certain community, s/he may put more effort into building better infrastructure in the area and, hence, may change the well-being of the households through other channels. Putting this aside, the authors conclude that fertiliser usage has a significantly positive impact on household well-being.

HOW TO EVALUATE AGRICULTURE PROJECTS?

Writing an impact evaluation plan can be a good starting point for any impact evaluation, and has the potential to save time and money. The impact evaluation plan should start with a description of the programme, which details its objectives and its design. The objectives of the programme will help the evaluator with building the results chain and selecting the performance indicators, whereas the design of the programme will determine which evaluation method(s) should be deployed in order to achieve a correct estimate of the programme's impact. The impact evaluation plan should then continue with detailing how the sample will be selected, what kind of data will be needed, and how and when the data will be collected. Finally, the expected outputs of the impact evaluation should be summarised, and the overall budget of the programme should be detailed.

A results chain starts with the inputs that are necessary for implementing of the programme. All these inputs, activities and outputs are aimed at reaching intended outcomes. Intermediate outcomes are the ones that are expected to be realised in the short run, and which are less in the control of the implementing agency. They may or may not be the

Figure 1: Building the results chain



Source: Gertler et al. (2010)

objective of the programme, but through the realisation of the intermediate outcomes, final outcomes are achieved and the programme serves its ultimate purpose.

The World Bank (2011) recently conducted a review of impact evaluations in agriculture, and came up with a list of the most commonly used performance indicators in such evaluations. The most common indicator used was the yield, defined as production or labour per total area of cultivated land. Of the studies reviewed in this report, 39 per cent used yield as their performance indicator. Income, defined as earnings from all activities, was the next most used at 24 per cent. Nine per cent of the reviewed studies used production and 8 per cent used profit as their performance indicator.

While building the results chain, the evaluator should start thinking about how long the implementation phase will last and how long it will take to produce the outputs and observe the intermediate and final outcomes. An evaluation done too early may conclude that the programme was ineffective although it turns out not to have been. Similarly, an evaluation done too late may attribute the impact of outside factors to the programme, producing biased estimates. If the budget permits, it might be a good idea to do an interim evaluation just when the intermediate outcomes are realised, and a final evaluation after the final outcomes are actually observed. The interim evaluation will provide early feedback, and will allow the implementer to redesign the programme before it is too late. The final evaluation, then, will determine whether the programme achieved its ultimate purpose.

The second step is choosing the methodology or design of the evaluation. The design of the programme itself will lead us to the right design for the evaluation. Is participation assigned randomly? Are there significant differences between eligibility and actual treatment? Are there confounding factors that affect both participation and the outcome variable? At which level is participation or actual treatment assigned? Are the outcome variables quantifiable? These are all questions that need to be answered in order to decide which of the evaluation methods discussed above should be deployed. Table 1 summarises which impact evaluation technique should be deployed in each case.

The third step is choosing the sample and testing the hypothesis. If there were no budget and time constraints, the ideal scenario would be to collect data on the whole population rather than a sample of it. However, if the population is too large this would not be feasible. A careful sampling should be done to arrive at an accurate and precise estimate of the programme's true impact. Once it is decided which performance indicators will be used and how large the sample should be to ensure accuracy and precision, the next step is to collect data.

Table 1: Picking the right methodology

Intervention type	Suitable evaluation approach
A lottery is done and farmers are randomly selected into the programme	Randomised controlled trial
Farmers in 14 districts are selected into the programme and the rest are left out	Difference in differences
Farmers from villages where a member of Parliament lives are selected into the programme	Instrumental variables estimation

In order to build a valid counterfactual, data should be collected both before and after the implementation of the programme. Data collected before implementation are called baseline data. Baseline data are crucial for determining whether the control and treatment groups were similar pre-intervention. Data collected after implementation are called post-intervention data and are used to measure the programme's impact along with the baseline data. More rounds of data collection can be helpful, especially if one wants to distinguish between the short- and long-run impacts of the programme.

Finally, two types of evaluation report are generally produced for an impact assessment project. The first is the baseline report, which is produced earlier in the evaluation timeline. This report aims to give an overview of the evaluation plan and to provide some summary statistics using baseline data. The second type of evaluation report is the final evaluation report. This is the most important written output of an evaluation project and

needs to be dealt with extremely carefully. It is as important to effectively communicate the results of an impact evaluation as it is to identify the programme's true impact. In addition to the comprehensive evaluation report, one- or two-page policy briefs may be produced to provide a quick snapshot of the evaluation. This can be especially useful in communicating the results to the general public in an efficient way.

INSTITUTIONALISING IMPACT EVALUATION

The World Bank (2009) has defined a framework for building an institutional setting for impact evaluations. The whole process should be country-led (i.e. it should be managed by the central government or a major sectoral agency). There should be firm acceptance by all stakeholders, including policy-makers, politicians, agencies and budget planners, and the system needs to be depoliticised so that it is not affected by changes in the administration or the government. Sound legislation and administrative directives will lead to the robust operation of the system. The stakeholders should be provided with clear guidelines on procedures and the methodologies.

Impact evaluations should be well integrated with the monitoring and evaluation systems and the budget and development planning activities. The independence, openness and accountability of the impact evaluations are extremely important. Independence should be guaranteed and protected by law, and the government should be open to evaluation findings and ready to accept and publish evaluation findings (including those it does not like). Data should be made available to the general public for further analysis. This institutional setup should also take a leading role in capacity development. There should be a supply of technical expertise to collect, analyse and conduct impact evaluations, and to create a strong demand for impact evaluations.

Although the above serves as a guide to setting up the institutional framework for conducting impact evaluations, there is no single best way to do it. The institutional setting should be compatible with political and public administration systems, and consistent with national impact evaluation capacity.

CONCLUSION

This briefing paper has outlined the main quantitative impact evaluation methodologies that can be applied to agricultural projects. Each technique has been illustrated with examples from actual impact evaluations in the agricultural sector of various African countries. The paper has also presented practical ways in which impact evaluations can be applied and institutionalised. The policy brief is intended to provide a synopsis of these tools for policy and budget practitioners who desire to have an intuitive understanding of impact evaluations in the agricultural sector. These tools were presented and discussed at CABRI's second Agriculture Dialogue in April, 2014, where finance, agriculture, and

monitoring and evaluation officials from 13 countries met to engage in a dialogue on the use of evaluation tools to enhance value for money. For more information on the dialogue and for additional background documentation, visit CABRI's website (www.cabri-sbo.org).

REFERENCES

- Baker JL (2000) *Evaluating the impact of development projects on poverty. A handbook for practitioners*. Washington DC: World Bank.
- Deininger K, Ali D & Alemu T (2009) *Impacts of land certification on tenure security, investment, and land markets: Evidence from Ethiopia*. Environment for Development Discussion Paper Series. Available at: <https://openknowledge.worldbank.org/bitstream/handle/10986/6897/WPS4764.txt?sequence=2> [accessed 24 March 2014].
- Duflo E, Kremer M & Robinson J (2008) *How high are rates of return to fertilizer? Evidence from field experiments in Kenya*. American Economic Association Meetings, New Orleans, January. Available at: <http://economics.mit.edu/files/3751> [accessed 24 March 2014].
- Gertler P, Martinez S, Premand P, Rawlings L & Vermeersch C (2010) *Impact evaluation in practice*. Washington DC: World Bank.
- Kaminski J (2011) Cotton dependence in Burkina Faso: Constraints and opportunities for balanced growth. In Chuhan-Pole P & Angwafo M (eds) *Yes Africa can: Success Stories from a dynamic continent*. Washington DC: World Bank.
- Kijima Y, Sserunkuuma D & Otsuka K (2006) How revolutionary is the NERICA revolution? Evidence from Uganda. *The Developing Economies* 44(2): 252–267.
- Kijima Y, Otsuka K & Sserunkuuma D (2008) Assessing the impact of NERICA on income and poverty in central and western Uganda. *Agricultural Economics* 38(3): 327–337.
- Ricker-Gilbert J & Jayne TS (2010) *What are the dynamic effects of fertilizer subsidies on household well being? Evidence from Malawi*. Available at: <http://ageconsearch.umn.edu/bitstream/96650/2/131.%20Dynamic%20food%20security%20effects%20in%20Malawi.pdf> [accessed 24 March 2014].
- World Bank (2009) *Institutionalizing impact evaluation within the framework of a monitoring and evaluation system*. Available at: https://ieg.worldbankgroup.org/Data/reports/inst_ie_framework_me.pdf [accessed 24 March 2014].
- World Bank (2011) *Impact evaluations in agriculture: An assessment of the evidence*. Available at: https://ieg.worldbankgroup.org/Data/reports/ie_for_ag.pdf [accessed 24 March 2014].

ACKNOWLEDGEMENT

This briefing paper is based on the Keynote Paper on Tools to Assess Value for Money in Agriculture Spending, produced by Development Analytics for the CABRI Agriculture Dialogue in April 2014.



CONNECT • SHARE • REFORM

For information on the Collaborative Africa Budget Reform Initiative, or to obtain copies of this publication, please contact:
CABRI Secretariat, PostNet Suite 314, Private Bag X 06, Waterkloof, Pretoria 0145, South Africa
E-mail: info@cabri-sbo.org
www.cabri-sbo.org