

# Meta-Evaluation of Extension Evaluation Case Studies



**Barry Pound, Sabine Gündel,  
Adrienne Martin and Essie Apenteng**  
(using a matrix developed by Ian Christoplos)

**2011**

GFRAS is the Global Forum for Rural Advisory Services. The GFRAS forum is made up of various stakeholders worldwide who have an interest and role in rural advisory services (RAS). The mission of this forum is to provide advocacy and leadership on pluralistic, demand-driven rural and agricultural advisory services within the global development agenda. GFRAS advocates for better recognition of the essential role of RAS and to see this recognition reflected in the policies of international development organisations, better investment in RAS, and more influence by RAS providers in international fora. The GFRAS vision is to see rural advisory services effectively contributing to the sustainable reduction of hunger and poverty worldwide.

Published by:

GFRAS – Global Forum for Rural Advisory Services  
c/o Agridea  
Eschikon 28  
8315 Lindau  
Switzerland  
Tel: +41 (0)52 354 97 64  
Info@g-fras.org  
www.g-fras.org

Text contributors

Barry Pound<sup>1</sup>  
Sabine Gündel  
Adrienne Martin  
Essie Apenteng

Coverpage Photo: Barry Pound,  
Natural Resources Institute

Design & Layout:

Ana María Hintermann-Villamil,  
webhint.ch, Gümüli, Switzerland

2011, Lindau, Switzerland

---

<sup>1</sup>Lead author: Natural Resources Institute, University of Greenwich,  
Chatham, Kent ME44TB, UK. B.pound@gre.ac.uk

## Introduction

The Global Forum for Rural Advisory Services (GFRAS) has commissioned the Natural Resources Institute to develop a toolkit for the evaluation of extension (projects, programmes, tools and initiatives). This commission has a number of components:

- A meta-evaluation of 15-20 evaluation case studies (presented here)
- A meta-review of the literature relevant to extension evaluation methods
- A workshop with practitioners and experienced evaluators to discuss the findings of a) and b) and to identify an initial set of tools
- A proposal for testing the proposed tools in a second phase of the project
- A brief of the toolkit for policymakers.

The overall purpose of this project is to identify methods for better evaluation of extension through the development of a toolkit for extension evaluation. The meta-evaluation and meta-review will also provide an in-depth basis for the selection of the approaches, methods and tools in the toolkit.

The mission of the Global Forum for Rural Advisory Services (GFRAS) is *"to provide advocacy and leadership on pluralistic, demand-driven rural advisory services (RAS). GFRAS does this in the context of the global development agenda, with a goal of promoting sustainable growth and reducing poverty"*. Its emphasis is on agricultural development and it defines RAS in this con-

text as "all the different activities that provide the information and services needed and demanded by farmers and other actors in rural settings to assist them in developing their own technical, organisational, and management skills and practices so as to improve their livelihoods and well-being".

## The Meta-Evaluation of Studies

The meta-evaluation presented here examined 17 case studies (Table 2) from a sample of 58 documents, chosen against a set of 4 criteria (Table 1). This ensured that the case studies represent a wide range of extension situations, scales, evaluation purposes and methods. The purpose of the meta-evaluation is to highlight current learning from extension, and to contrast the quality and range of evaluation approaches. The meta-evaluation does not question the point and use of extension in general, nor does it make value judgements about the projects evaluated. However, it does assess the extent to which the case studies questioned the role and position of extension in their national development systems.

In combination with the meta-review of extension evaluation literature, the evaluation is intended to be a resource for use in developing a toolkit for extension evaluation.

**Table 1: Criteria used for the selection of Case Studies**

<b>Criterion</b>	<b>Explanation</b>
1. Type of extension system/situation	Innovative, pluralistic, demand-driven, multi-stakeholder and cutting-edge situations – involving a range of stakeholders. This should ensure that the tools used for evaluation are suitable for such systems/situations.
2. Scale of extension evaluation	A range of scales from national and institutional level to programme and project level. The scale will influence the range of suitable tools.
3. Types and purpose of evaluation	Ex-ante, ex-post, process- use, use of findings, horizontal evaluation, vertical evaluation.
4. Methods and tools used for evaluation	Range of different evaluation methods and tools (e.g. rapid evaluation and assessment methods (REAM)), real-time evaluation (RTE), participatory evaluation, quantitative and qualitative methods, surveys, etc.

Anyone interested in the theoretical background of some of the issues raised by this meta-evaluation should consult the review of literature relevant to extension evaluation<sup>2</sup>.

All of the cases chosen (Table 2) are taken from the last ten years (fourteen from the last five years), and from a range of commissioning (bi-lateral and multi-lateral donors, NGOs and national extension

systems) and implementing organisations in Asia (five), Africa (nine, some covering several countries), the Far East (one country), Europe (one study in six countries) and one multi-region study.

The detailed interrogation for each case study was carried out using a matrix of questions drafted by Ian Christoplos and further developed by Adrienne Martin and Barry Pound (Annex One).

---

<sup>2</sup>Adrienne Martin, Sabine Gündel, Essie Apenteng, and Barry Pound, *Review of Literature on Evaluation Methods Relevant to Extension* (Chatham: Natural Resources Institute, 2010).

**Table 2: Details of the 17 case studies and their assessment against the 4 selection criteria**

No	Author and year	Source document title
1	Sanne Chipeta, 2008	Extension as a tool for farming as a business – Agricultural Support Programme, Zambia
2	MAAIF, 2005	Proceedings of the MTR of NAADS – Uganda
3	Raul Hopkins, 2000	Impact Assessment of OXFAM Fair trade
4	Lorenz Bachmann, 2006	Ten years work of Misereor partners on sustainable and organic agricultural practices – Uganda
5	SDC, 2006	Joint External Review (SDC/WB) Rural Advisory Service and Kyrgyz Swiss Agricultural Program
6	Willem van Weperen, 2003	Impact Assessment Study of Southern Highlands Dairy Development Project – Tanzania
8	Kristin Davis (2010)	In-Depth Assessment of the Public Agricultural Extension System of Ethiopia
9	Kristin Davis, 2010	Impact of Farmer Field Schools on Agricultural Productivity and Poverty in East Africa
10	Catherine Laurent, 2006	Agricultural Extension Services and Market Regulation – EU

Selection criteria			
Type of extension situation	Scale of extension evaluation	Purpose of evaluation	Methods used for evaluation
Demand-driven, multi-stakeholder project	National	End of phase review	Appreciative Inquiry
Innovative, pluralistic, demand driven national system	National	Mid-term review	Questionnaires, Interviews, Stakeholder workshop
Innovative, NGO	Multi-region	Periodic review	Economic Impact Ratio, Video
Multi-stakeholder, NGO	National	Periodic review	Questionnaire questions Pre- and post-survey workshops
Long-term support to national programme	National	Planning review	Interviews, Field visits
Long-term project, multi-stakeholder	Sub-national	Final mission report	Interviews
National programme	National	Planning review	Key informant interviews, FGD, Field visits
Demand-driven, innovative method	Regional	Longitudinal impact study	Semi-structured interviews
Multi-stakeholder, extension regulation	Regional	Academic	Review of documents, institutional analysis, organizational

<sup>3</sup>The ASP review was not a stand-alone activity, but an analysis of the extension component of the ASP - which was supplemented by other studies, such as impact studies and a cost-benefit analysis.

**Table 2: Details of the 17 case studies and their assessment against the 4 selection criteria (cont.)**

No	Author and year	Source document title
12	Ian Christoplos, 2010	Report of the External Review Mission for the Laos Extension for Agriculture Project
13	Melinda Cuellar 2006	The National Agriculture and Livestock Extension
14	DANIDA, 2004	Farm Women in Development – India
15	Paul van Mele, 2007	Videos that Strengthen Rural Women’s Capability to Innovate – Bangladesh
16	Steven Franzel, 2009	Assessing the Costs and Benefits of a Volunteer Farmer Trainer Programme in Meru, Kenya
17	Frank Place, 2008	The impact of fodder trees on milk production and income among smallholder dairy farmers in East Africa



<b>Selection criteria</b>				
<b>Type of extension situation</b>	<b>Scale of extension evaluation</b>	<b>Purpose of evaluation</b>	<b>Methods used for evaluation</b>	
National system	National	End of phase review	Key informant interviews	
National system	National	End of phase evaluation	Interviews, Focus group discussions, SWOT analysis, Participatory force-field analysis	
Innovative, multi-stakeholder	National	Longterm support study	Analysis of project data Stakeholder interviews, Questionnaires, Key informant interviews Knowledge, Attitude and Practice (KAP)	
Innovative intervention	National	Academic	Questionnaire, KAP surveys	
Innovative intervention, demand driven	Sub-national	Interim assessment	Cost-benefit evaluation, Interviews, Gross Margin Analysis, Internal Rate of Return	
Innovative, demand-driven	Regional	Academic	Community workshops, PRA, Focus group discussions Case studies Interviews with key informants	

# Findings of the Meta-Evaluation of Extension Evaluation Case Studies

## Scale, duration and resources

The 17 case studies vary greatly in terms of their **scale** (from evaluations of multi-regional programmes such as the Oxfam Fair Trade study (#3, Table 2) and multi-country studies such as the comparison of market regulation of extension in six EU countries (#10, Table 2) to very local studies, such as one on volunteer farmer trainers in Meru District of Kenya (#16, Table 2). This sample also demonstrates the wide **range of subjects** covered by the evaluations chosen as case studies, which also includes evaluation of national extension systems (e.g. in Indonesia, #7 and Ethiopia, #8) and of specific tools, such as Farmer Field Schools (#9) and video (#15).

Just as striking is the range of **duration** and **resources** committed to the evaluations. This range extends from a single consultant working for 10–15 days to the DANIDA study of Farm Women in Development

(#14), which lasted for 3 months and used 8 experts and 10 local technicians. A more typical study was one involving around 3–4 consultants working for 2–3 weeks, representing a cost of US\$50–100,000 with living allowances, flights, local enumerators and internal transport costs. While considerable, this investment constituted a small percentage of the overall project or programme cost (>US\$100 million for the first 7-year phase of NAADS in Uganda (#2) and US\$23.5million for World Bank support to extension services in Indonesia (#7)). It was also noted that some projects have been running for a long time without a comprehensive evaluation of their performance – 15 years in the case of Farm Women in Development in India (#14). Only one study cited limited time as a major constraint to evaluation. This was the study of the Laos Extension for Agriculture Project (#12), in which only a week was available for fieldwork with farmers, curtailing the use of rigorous fieldwork methods. It is suspected that in many cases constraints of time and human resources dictated the methods and the size of the sample used. Section six of a sister document which re-

---

<sup>4</sup>Adrienne Martin, Sabine Gündel, Essie Apenteng, and Barry Pound, *Review of Literature on Evaluation Methods Relevant to Extension* (Chatham: Natural Resources Institute, 2010).

<sup>5</sup>OECD, *DAC Principles for Evaluation of Development Assistance* (Paris: Development Assistance Committee, 1991).

views the literature relevant to extension evaluation examines donor standards for evaluation. The document setting out the DAC principles (OECD, 1991<sup>5</sup>) includes a design standard relating to resources, as follows: "**Design and implementation of evaluation** – terms of reference for the evaluation should define its purpose and scope and the intended recipients or users. It should include the questions to be addressed in the evaluation, specification of the methods, the performance assessment standards and the resources and time allocated".

**Observation: Evaluation resources to be commensurate with project value and sufficient to allow the use of methods appropriate to the objectives and uses of the evaluation.**

### Project documentation

A major surprise to the authors of this study was the almost complete absence of project or programme results-based frameworks (e.g. logical frameworks) against which projects and programmes could be evaluated. An exception is the 2005 Mid-term Review of the National Agricultural Advisory Services Programme in Uganda (NAADS), which refers back to the Programme logical framework that was developed in 2000<sup>6</sup>. The Southern Highlands Dairy Development project had project frameworks, but decided against

trying to evaluate performance against them as it would have meant evaluating against 7 different frameworks, one for each of the seven SHDDP phases. This was not realistic for a comparatively low-budget evaluation (Willem van Weperen, personal communication).

The lack of results-based frameworks detailing expected project outputs, activities and time-bound quantitative and qualitative indicators makes it difficult for an evaluation team to assess whether a project is meeting the expectations set at the start of the project or not.

Several evaluations started their work by reading and analysing project documents, including quarterly and annual reports to their funding agencies. In cases where they cite quantitative progress against targets and milestones (or logical framework activities and indicators) these project reports can be very helpful in charting the course and achievements of the project.

**Record keeping** by the beneficiaries can also be very useful to the evaluation team. For example, FARM-Africa trainers (one of the groups being evaluated in #16) kept records of their costs and revenues. These were useful in calculating cost-benefit ratios and one aspect of the viability of the intervention.

---

<sup>6</sup>NAADS Master Document of the NAADS Task Force and Joint Donor Groups, MAAIF, October 2000.

**Observation: Evaluation starts at the project design stage – results-based frameworks such as logical frameworks, reporting procedures and formats, internal and external M&E budgetary provision and simple record keeping by beneficiaries.**

### Clear objectives, uses and users of the evaluation

The **objectives** are clearly stated in most case studies. However, the uses to which the report will be put and the anticipated users are often not specified. This is a serious omission, as the style, level and content of the report should be **tailored** to such uses and users. In some cases the users might be governments or national organisations (e.g. the governments of Kyrgyzstan and Indonesia) whose first language is not English, and yet all the reports were in English and did not even contain an abstract in the local language. In no case were the users **differentiated** and recommendations given for each separate user. There should be a clear link between the objectives and uses of the evaluation and the **methods** used to develop and apply the sampling framework (locations, actors, households and individuals), and to analyse, confirm and present the findings.

### Project/programme logic

While most of the evaluations describe what the project does, few of them clearly state the **project logic** (if this is done, **then** that will result). The International Finance Corporation<sup>7</sup> suggests that this is an essential step in evaluation methodology as it provides a basis for evaluation (has **this** been done, has **that** resulted?).

One study that used the existing project logic as a starting point was the evaluation of the Agricultural Services Project in Zambia (#1, Table 2). Figure 1 (page 12, questions on page 13), shows a diagram of the logic for that project (on the left), while on the right are the main evaluation questions used to assess the project's performance and impact. While there is a good correlation between the two (e.g. in the areas of markets, productivity, farmer organisation, quality control, and finance), there are still some questions that do not correspond to a component in the logic diagram (e.g. policy influence and pluralism).

---

<sup>7</sup>Guidelines for Monitoring and Evaluation of Agribusiness Projects, IFC Advisory Services, International Finance Corporation, June 25, 2009.

**Observation: A statement of the project logic can provide a clear position on the assumed input:outcome relationship of the project. The evaluation should test whether that expected relationship has been realised in practice.**

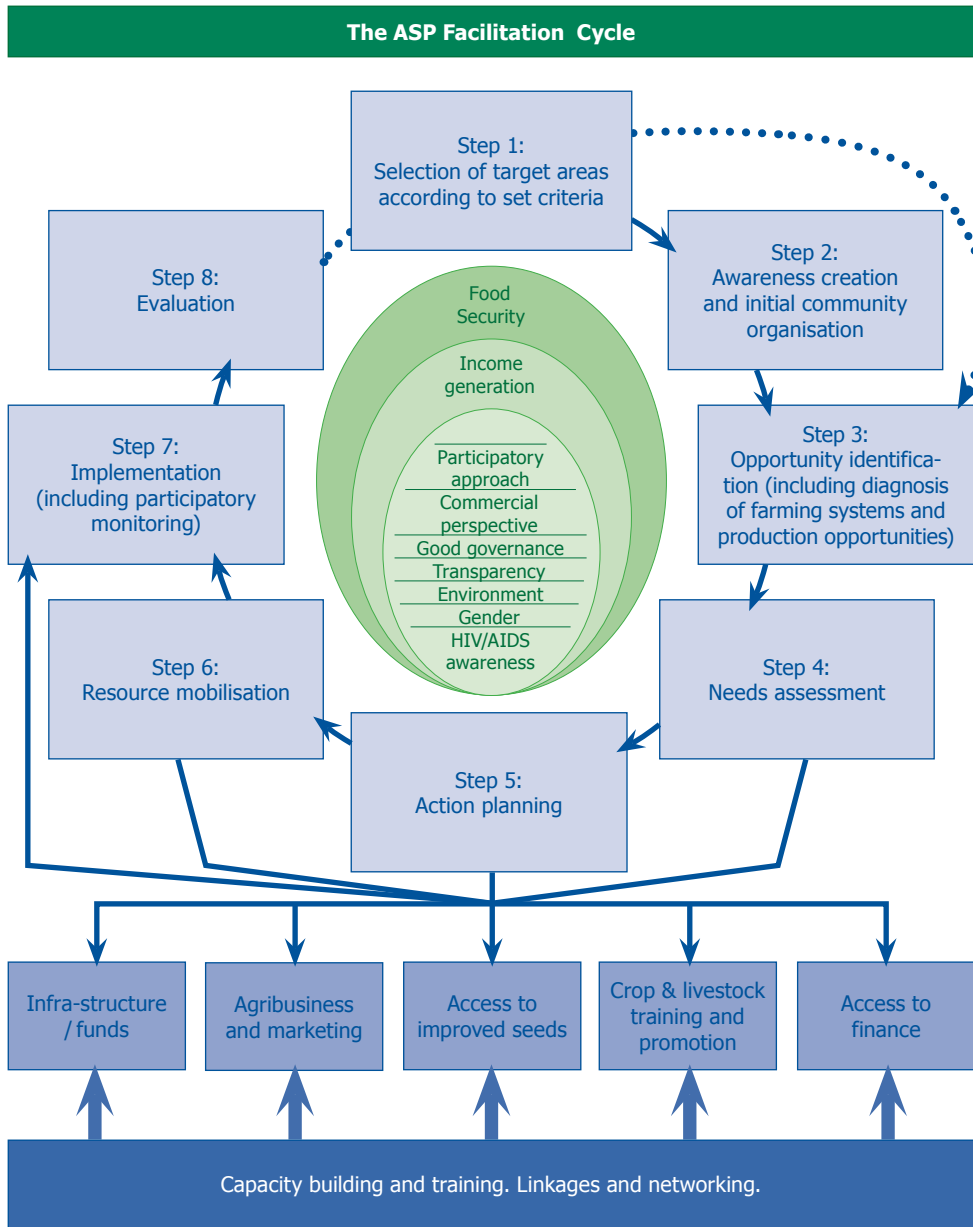
### Context and linkage to the bigger picture

Any project or programme is implemented within a local, national and international context. The main contextual facets are those of **policy** that enables or limits the effectiveness of the project, **institutions** and their capacities and linkages, **infrastructure** that supports or constrains effectiveness, and **markets** (inputs and sales) that push, pull or stifle enterprise. International policies – for instance on global public goods and international trade that encourages exports – can radically influence the targeting and content of extension programmes.

Oxfam's Fair Trade (#3) clearly depends on a particular type of export market, but it also depends on government attitudes towards labour rights and adequate infrastructure at ports and airports. Thus it is appropriate to evaluate the achievements of the projects against these con-

textual issues. Other case studies have policy at their heart, such as the evaluation of the impact of market regulation on extension services in six EU countries (#10). In general, the case studies have done a reasonable job of **describing** the national context in which they operate, while not **analysing** the strengths and weaknesses of this in relation to the project. An exception was the Joint External Review of the Rural Advisory Service and Kyrgyz Swiss Agricultural Program, #5, which had an explicit criterion to *analyze the contextual changes in the agricultural sector in Kyrgyzstan*. This it did. For example, the policy context is summarised in the Box below. This information was taken into account when making recommendations for the future funding and direction of the programme to increase coverage; reduce costs; improve cost recovery from farmers; reorganize governance to give management greater control over key decisions; improve organizational arrangements for client (farmers) representation; improve responsiveness to clients (farmers); improve the mandate and payment system to increase performance incentives; and improve staff recruitment, payment systems and staff training to increase performance incentives and raise the technical quality of services.

Figure 1: Agricultural Services Project components and consequent results (Chipeta et al, 2008) compared to the evaluation questions used to assess its performance and impact



## ASP evaluation questions

1. How does policy support the aims of the ASP; how does ASP support policy development?
2. How does extension improve farmers' linkages and voice in the market and knowledge system?
3. How does extension support farmers' ability to increase productivity?
4. How does extension improve farmers' ability to solve their own problems?
5. How are farmers and their organisations involved in defining the content of extension?
6. What are the quality issues in extension?
7. How does the ASP increase farmers' access to markets?
8. What is the relation to private sector actors?
9. How is the ASP extension financed; how sustainable is this?
10. To what extent has ASP been institutionalised and decentralised, and how is pluralism promoted?

### Box 1: Policy context summary: Joint External Review of the Rural Advisory Service and Kyrgyz Swiss Agricultural Program

The main programme policy considerations are whether the RAS can be considered a public agricultural advisory service, and if so what proportion of the agricultural population it should serve (at the moment it only serves about 5%), the funding it should get from government and from farmers (public vs private goods), and whether there is a continuing case for donor support.

Other policy areas include:

- decentralisation of services
- autonomy of the program from donors
- increasing demand drive (the mandate system)
- financial sustainability
- improved representation of farmers
- the need for external monitoring

Nationally the RAS programme responds to the privatisation of land in the mid-1990s and the creation of many small private farms, but with low levels of skills in private farming.

There is, however, very little information in the cases on how projects/programmes relate to **international issues** (global trade, global funding opportunities, climate change, carbon trading etc.) or (apart from the evaluation of the Ethiopian extension system – #8) any comparison of the project under review with extension initiatives elsewhere in the world.

Context is not static, as liberalisation and decentralisation in countries like Ethiopia have shown over the last 20 years. It is therefore also important to signpost the **direction** in which the most significant aspects of context are moving, especially for those evaluations looking to make recommendations for future project phases.

**Observation: The performance of a project, programme or tool is bound by its operational context. It therefore essential for the evaluation to acknowledge that context and to highlight its strengths, weaknesses and trends and their implications for future programme design.**

### Appropriate evaluation criteria

Thirty percent of the studies set out their evaluation criteria (or indicators or hypotheses) clearly, so that it is easy for the reader to understand exactly what was evaluated.

For example, the staff and evaluation consultants for the Tanzania Southern Highland Dairy Development Project (#6) developed three **hypotheses** which described the project's assumptions:

- Milk production has made a positive contribution to the farming system of the Southern Highlands
- The position of smallholder dairy farmers has been strengthened by building farmer organizations
- The project was a relevant and successful actor in the development of the dairy sector (and - in a broader way – of crop-livestock based agriculture) in the Southern Highlands

They then formulated a **set of questions** to ask stakeholders in order to prove or disprove these hypotheses. This set of questions formed the basis of their evaluation of the project.

Perhaps the most impressive example is the World Bank evaluation of extension in Indonesia (#7), which developed a set of 39 **indicators** (covering income, welfare, productivity, technology, information, inputs, markets, capital, empowerment, participation, gender and poverty) for which they gathered qualitative and quantitative information. Another clear example is the evaluation of the Agricultural Services Project in Zambia, which assessed achievements against ten **perspectives**, as shown in Figure 1.



About half of the cases referred to the five **OECD DAC criteria** – relevance, efficiency, effectiveness, impact and sustainability. One project (Evaluation of LEAP, Laos – #12) uses the DAC criteria in its main evaluation questions: Is LEAP contributing to the effectiveness and sustainability of NAFES? Is LEAP moving into livelihoods? Is LEAP becoming institutionalized? Are LEAP's outputs useful and used? Is capitalisation moving LEAP beyond just another project?

In its Strategic Findings section, the report then specifically deals with **Relevance** of LEAP goals and strategy; **Feasibility** in light of overall findings; **Effectiveness** in achieving outputs and outcomes; **Efficiency** in relation to partnerships and **Sustainability**.

However, it must be remembered that the DAC criteria are **generic**, so there is still the need to develop specific criteria for the individual situation of the subject under evaluation.

**Observations: a) after clear objectives, clear evaluation questions are the next step in successful evaluation; b) there are different ways of formulating the evaluation questions, but they should ultimately answer the questions asked by the ToRs; c) it may be that new issues are uncovered during the evaluation. These would then need to be incorporated into the evaluation questions (i.e. retain flexibility).**

## Attribution

As Franzel (#16) observes, "*Evaluating an extension approach is always problematic because a single extension method, e.g., farmer trainers, is never independent of other methods, e.g., radio shows, flyers, posters, etc.*". Similarly, an extension project is rarely independent of other initiatives and events. For example, the impact of drought was greater than the impact of project interventions on yield/income in the Farm Women in Development project in India (#14).

Six of the 17 cases studied (#2, 4, 7, 9, 15, 17) included a counterfactual to attempt to isolate the effects that were due to the project intervention alone. One of these (#7) had a counterfactual in which three types of farmers were surveyed – those not supported by the Misereor partners (the control or counterfactual group), those who had received support for 2–5 years (short-term adopters), and those who had received support for 6–10 years (long-term adopters). Another project (Extension in Indonesia, #7) also used three samples of households: 450 project households (from a population of 30,000), 450 non-project households in a different District, and "spill-over" households within the project Districts. This allowed comparisons between with- and without-project situations, and the spill-over effects of the project to be assessed. The construction of a counterfactual is good discipline in that it forces the evaluator to define what the

project is supposed to be achieving, and what might not be achieved without the project. It therefore expresses the project logic (whether explicit or not) and underlying hypotheses about what aspects of change the programmes should be able to impact on and what they will not.

## Stakeholders

None of the case studies reported a **stakeholder analysis** as part of their evaluations (although the Evaluation of Agricultural Extension Services and Market Regulation in EU countries (#10) conducted institutional and organisational analyses to understand the evolution of institutional configurations and their reaction to market regulation trends). Such an analysis can identify those stakeholders that are important (positive and negative influences) to achieving project objectives. This helps to prioritise those that should be included in samples or interview schedules because of their particular perspectives. It may also be that such an analysis can identify those with whom the evaluation findings should be confirmed, or to whom they should be disseminated. Without such an analysis there is the risk that the evaluation will be project-centric, ignoring the role and influence of national-level actors or the private sector. An example is the evaluation of Misereor extension interventions in Uganda (#4), which concentrates almost exclusively on the beneficiary farmers and a

counterfactual set of farmers. This can be contrasted with the Mid-Term Evaluation of NAADS in Uganda (#2), which elicited the views of a very wide set of actors through a diverse set of evaluation activities, including a tour of two Districts by Members of Parliament. The majority of the evaluations did not list which stakeholders they included, so an analysis of the inclusion/exclusion of particular perspectives was not possible.

## Capacity development

Only one study (ASP Zambia, #1) had an explicit objective to develop the capacity of local staff in evaluation. However, the majority of studies involved some local staff – either as members of the core technical evaluation team or as enumerators. In addition, the involvement of local stakeholders in the planning and feedback workshops will also have provided some level of capacity building. As detailed in the review of literature relevant to extension evaluation<sup>8</sup>, several of the donor standards emphasise capacity development (e.g. OECD<sup>9</sup>, IDRC<sup>10</sup>).

<sup>8</sup>Adrienne Martin, Sabine Gündel, Essie Apenteng, and Barry Pound, *Review of Literature on Evaluation Methods Relevant to Extension* (Chatham: Natural Resources Institute, 2010).

<sup>9</sup>OECD, *Quality Standards for Development Evaluation, DAC Guideline and Reference Series* (Paris: OECD, 2010), <http://www.oecd.org/dataoecd/55/0/44798177.pdf> (accessed 6 June 2011).

<sup>10</sup>IDRC, *Guiding Principles of IDRC's Evaluation Unit* (Ottawa: International Development Research Centre, n.d.), [http://www.idrc.ca/uploads/user-S/12095810441/Evaluation\\_Unit\\_Guiding\\_Principles.pdf](http://www.idrc.ca/uploads/user-S/12095810441/Evaluation_Unit_Guiding_Principles.pdf) (accessed 21 September 2011).

Evaluations may need to provide time in their schedules for the training of enumerators, facilitators and data entry technicians in order to ensure that competent and accurate information is obtained and recorded.

## Reporting the evaluation findings

### a) Feedback to local stakeholders and commissioning organisations

Most of the studies feed back their findings at least to the project staff, and others go much further in inviting a range of stakeholders (including farmers in a few cases). The NAADS Mid Term Review (#2) held a 2-day review of the findings from four different evaluation components with 200 participants drawn from a wide range of participants. Similarly, the review of the extension system in Ethiopia (#8), held a stakeholder meeting for extension personnel, researchers, NGOs, and policymakers to validate and refine the findings and recommendations for the final version of the document. A two-and-a-half-day workshop with DAs, SMSs, regional bureau heads, extension heads, MOARD staff, research staff, ATVET heads, and Sasakawa-Global 2000 staff was held to brief them on the findings and obtain feedback and validation, and to discuss in detail with these stakeholders how to actually implement the recommendations.

### b) Presentation and dissemination of evaluation findings

Not all reports included an executive summary, and in no case (where there was more than one named user) were the recommendations divided for the attention of each of these main users. A few reports (ASP Zambia #1, SHDDP Tanzania #6, Agricultural Extension in Indonesia #7, Agricultural extension and market regulation #10 and Use of video in Bangladesh #15) had no recommendations whatsoever. In only one case (#8) were the recommendations ranked to give the user some guidance on their weight. Some reports used tables, diagrams and boxes attractively but there were few photographs of the extension interventions or the evaluation team activities.

All the reports are in English without any summaries in any local language, even where the national government or the project is an identified user of the findings and their first/official language is not English.

Several studies make an effort to disseminate their findings; the ASP (#1) and NAADS project (#2) via the Neuchatel Group and NAADS websites respectively, OXFAM Fair Trade (#3) through the use of videos, and the Bangladesh project to evaluate the use of video to Strengthen Rural Women's Capability to Innovate by the distribution of CDs.

## Methods used in the Case Studies

The 17 case studies reported in this meta-analysis were selected partly for their use of contrasting methods (Table 2). However, two of the studies (NAADS Uganda, #2 and the Kyrgyz agricultural programme, #5 did not have a methods section in the report).

Some evaluations are dominated by a single qualitative method (Appreciative Inquiry and Knowledge, Attitudes and Practice in the evaluation of the Agricultural Services Project and the evaluation of video in Bangladesh, respectively). Others (e.g. the World Bank evaluation of extension services in Indonesia and the evaluation of Misereor activities in Uganda) are dominated by quantitative instruments (questionnaires). In contrast the NAADS evaluation has four components, each of which used a different set

of methods, the results from which were all discussed in a large stakeholder workshop – itself a potentially powerful evaluation tool. The Table in Annex 2 presents the methods used in each of the case studies, while Table 3, below, summarises the methods found in the overall sample:

A number of common themes run through this tabulation of methods (Annex 2 and Table 3):

- Only one project (NAADS Uganda, #2) reports the use of a project logframe or any other form of measuring progress against the original intended outputs of the project in a systematic way
- A small minority of studies (notably cases # 4, 9, 11 and 17) have a sampling framework for the selection of farmers or other stakeholders
- Similarly, a minority of evaluations uses a counterfactual (present in cases # 2, 7, 9, 15 and 17)

**Table 3: Methods used in the case studies**

Methods	Comment	Studies using
<b>Qualitative methods and tools</b>		
Appreciative inquiry	Non-threatening, participatory, qualitative method that can be used with different stakeholders.	#1
Knowledge, Attitude and Practice (KAP) surveys	Good for measuring changes over time	#14, 15
Focus group discussions	Can use socially differentiated groups to identify differences between them	#8, 13, 17
Semi-structured interviews/Key informant interviews	Checklists guide interviews, but allow a more natural conversation than questionnaires. More flexible, and effective when intelligent probing is used	#2, 5, 6, 8, 9, 10, 12, 13, 14, 16, 17
Participatory force-field analysis	Good for analysing future scenarios	#13
Institutional and organisational analysis	Rarely used, but especially useful for looking at sustainability	#10
Field visits / Observation	Should form part of the sampling framework and complement quantitative information. Checklists help to systematise information collection	#2, 5, 8
PRA	Wide set of tools that can be deployed to meet most situations	#17
Case studies	Can explore specific issues in depth and are more intimate than questionnaires	#17
Video	Useful visual record of a situation. Useful for dissemination of findings of evaluation.	#3
SWOT analysis	Strengths, weaknesses, opportunities and threats	#13

Methods	Comment	Studies using each method
<b>Quantitative</b>		
Questionnaire	Survey mainstay. Often overly complicated and long (but only 1 page in case of Bangladesh plant clinics evaluation)	#2, 4, 7, 10, 14
Economic impact ratio	Compares economic benefits with opportunity cost of labour (surrogate counterfactual)	#3
Cost-benefit analysis	Can do this for economic elements and then add qualitative costs and benefits into discussion	#16
Gross margin analysis	Useful for assessing impact of discreet interventions or enterprises	#16
Internal rate of return	Useful for assessing the return to extension investment	#16
Record keeping by project staff and beneficiaries	The four methods above can draw on this data if available (ideally needs to be incorporated into project M&E at design stage)	#3, 14
Baseline surveys	Very useful if the right data are collected and can be followed up in a way that allows direct comparisons to be made	#2, 5, 7, 15

Complementary methods/practices		Studies using each practice
Review of project and secondary documentation	Especially useful if project has tracked progress against outputs (result areas), targets or milestones, and has kept records of costs and benefits that can be analysed	#5, 6, 8, 9, 10, 12, 13, 14, 17
Pre-survey planning workshop	Good for developing/refining methods and questions and for promoting local ownership and capacity	#4, 14
Post-survey stakeholder workshop	Useful for confirming findings, discussing future scenarios and instilling ownership of recommendations	#2, 4, 5, 6, 8, 12, 13, 17
Sampling framework	Systematic way of arriving at (and demonstrating) a representative and/or statistically viable sample of the overall evaluation population	#4, 9, 11, 17
Random, stratified and purposive selection of sample	Appropriate method depends on objective; purposive selection good when specific criteria or socially differentiated groups are involved	#9, 4, 7 (random) #9 (purposive) #17 (stratified)
Opportunistic selection of sample	Dubious statistically, but often the most pragmatic!	#11
Counterfactual	Way of showing that benefits can be attributed to the project's interventions	#2, 4, 7, 9, 15, 17
Use of hypotheses and project logic to formulate evaluation questions and indicators	Shows that the evaluation team understands what the project is trying to achieve and makes a direct link between this and the evaluation questions	#1, 6
Longitudinal studies	Many evaluations are snapshots. Longitudinal studies assess changes over time	#5, 9
Logical framework	Disciplined way of summarising what the project is meant to achieve (quantitatively and qualitatively) and how and when, against what assumptions. Evaluations can use it as a yardstick, but shouldn't follow it too rigidly.	#2

- Surprisingly few projects have a baseline against which to measure progress (cases # 2, 5, 7 and 15) – even projects that have been operating for a long time (e.g. the Farm Women in Development project in India have not invested in a baseline set of data in over 20 years). The Tanzania Southern Highlands Dairy Project did conduct a socio-economic Baseline at the beginning of the project which was regularly updated through the project’s Socio Economic Analyses System (SEAS). However, the team judged its information as not good enough and not useful for purposes of impact assessment and so it was not used during the evaluation<sup>11</sup>.
- However, some studies (e.g. Kyrgyzstan Rural Advisory Service and the East African study on Farmer Field Schools) were effectively longitudinal. In the first case the project has had two previous external reviews so that progress can be assessed over time
- Questionnaires are used in several cases, but these vary greatly in complexity. The evaluation of extension services in Indonesia gathered information about 39 indicators in its questionnaire, while the evaluation of plant clinics in Bangladesh used a 1-page questionnaire
- While random sampling may be considered the gold standard, it has both advantages and disadvantages. One study (Evaluation of Farmer Field Schools, #9) used **purposive sampling**, one used stratified sampling (Fodder trees in East

Africa, #17), while another (Plant clinics in Bangladesh, #11) admitted that farmers were selected “opportunistically”!

- Two studies (Misereor, #4 and Farm Women in Development, 14) held **pre-survey workshops** to plan their fieldwork and elaborate the questions to be used. Most studies used a **post-survey workshop** to present their findings and get feedback, although the breadth of stakeholders present varied. A few also held de-briefings for funders and government line agencies.

**Observations: The case studies present a variety of evaluation and analysis methods, both qualitative and quantitative. No one method is perfect for all situations, and often a carefully selected combination of methods best captures the qualitative and quantitative consequences of project interventions.**

---

<sup>11</sup>The evaluation report (page 11, conclusion 21) concludes that: “Project monitoring tools were inconsistent, overambitious and not always well adapted to the local situation. As a result, the project ended up collecting a lot of data which it was not able to use as intended, to serve as a basis for decision making. Participatory self assessment and SAF, introduced at a later stage into the project, proved to be much more effective monitoring tools”.



## Good Practice

A finding of this meta-evaluation that surprised the main author was that the **procedures** used in the implementation of

the evaluation were just as important as the qualitative and quantitative methods. Table 4 lists the areas of good practice identified in the case study analyses:

Example of good practice	Comment	Studies demonstrating instances of good practice
Project/programme has an M&E strategy which provides the evaluation team with useful data from project reports, project staff and project beneficiaries. The M&E strategy should include periodic internal and external reflection and evaluation events. This adds up to a "culture of evaluation" to be encouraged within the project.	Needs to be part of project design	#2, 16
Project/programme has results-based (e.g. logical) framework or other means of tracking progress against outputs, activities, targets, indicators and assumptions	Needs to be part of project design	#2
Clear ToRs for evaluations that are achievable given the resources and time allocated to the evaluation; ToRs are appended to the report	Often ToRs are over-ambitious and complex for the time and human resources allocated	#1

Example of good practice	Comment	Studies demonstrating instances of good practice
Inclusion of stakeholders in planning the evaluation, setting evaluation questions, implementing the evaluation, discussing the findings and finalising the conclusions, while maintaining the independence of the evaluation	Increases ownership and can avoid mistakes being made by consultants unfamiliar with cultural and technical issues	#1, 6
Clear, concise executive summary	For those too busy to read the whole report	#7, 14
Clear identification of objectives, users and uses of the evaluation	Often only the objectives are clearly stated, with little attempt to define use and users of the evaluation	#2, 5, 6, 9, 12
Analysis of the local, national and international contexts of the project/programme and interventions	Each level has its own influences on how the project performed in the review period, performs now and will perform in the future. In addition, the context will relate the project/programme to the overall development situation in the country and the main agricultural policy thrusts, such as poverty alleviation, food security, commercialisation, NRM and climate change.	#7, 14
Use of tables, boxes, diagrams, photographs to complement text in the report	Different users prefer different ways of having findings presented to them.	Little use of photos

Example of good practice	Comment	Studies demonstrating instances of good practice
Inclusion of the evaluation questions in the report	Report should clearly show how the evaluation questions are arrived at and what they are. Means that anyone repeating the exercise can use the same questions	#1, 6, 7
Use of criteria for the selection of villages, Districts (woredas) etc.	Provides a discipline (rigour) for their selection	None included in report
Use of rigour at all stages (planning, implementation, analysis, presentation and discussion, reporting)	Rigour can be equally applied to qualitative methods and quantitative methods. However, rigour such as developing representative sampling frameworks and interviewing a sufficient proportion of the sample population requires resources and time. Those funding the evaluation need to be prepared to send teams to the field for long enough to accumulate reliable evidence	#7
Capacity building of local institutions to conduct future evaluations through collaboration and training	Preferably all stages, including planning and analysis	#1
Involvement of different stakeholders in components of the evaluation	e.g. MPs in the evaluation of NAADS	#2
Future projection: where is the project headed and what are its future needs and strategy in view of present context and future trends? Is the project needed at all or would a different investment be a better option?	Recommendations should try to cover a broad set of issues that affect the project (financial sustainability, human resources, policy support, scope and scale...) or present an alternative to the project	#1, 2, 7, 8

Example of good practice	Comment	Studies demonstrating instances of good practice
Where there is more than one user, provide separate recommendations tailored to each main user of the evaluation (e.g. donor, government, project...). Recommendations could also be ranked in terms of priority to guide users	None of the evaluations provided tailored recommendations for different users. None provided any sort of action plan to guide the users in implementing the recommendations	None, but #1 did rank interventions in order of effectiveness and #8 prioritised recommendations across 3 time "horizons"
De-briefing of funding agency and relevant line government agency	Provides feedback to key stakeholders and from them to evaluation team	#1, 5, 12
Promotion of evaluation findings to a wider audience using media (video, newspapers...)	Adds value to the findings and can support the launch of a new phase	#2 (website), 3 (video), 15 (CD)
Field testing of methods (e.g. questionnaire)	This provides an opportunity for training the enumerators, to assess whether questions are appropriate and to assess how long it will take per interviewee.	#4, 8, 11
Differentiation of impact between different social groups (men and women, wealth groups, young and old, remote and accessible etc)	Enables the impact on different groups to be compared, and may provide insights into how to address the differences	#3 (partial), 9 (partial), 17 NB social differentiation is general poorly addressed
Stakeholder analysis	Not used in any of the cases analysed here, but a useful way of identifying those who have positive or negative influence on, or interest in, the project or the evaluation findings.	None

## Conclusions

The evaluation of 17 case studies drawn from different regions of the world, of different types, of varying scales and using different methods, concludes that:

- The matrix used in the interrogation of the case studies provided a useful framework and discipline for systematic comparative evaluation between case studies, but made each case study evaluation very long-winded;
  - The resources available for most evaluations are not sufficient to allow them to conduct the fully comprehensive set of activities outlined in the following bullet points;
  - While the case studies cumulatively covered all the elements of a fully comprehensive evaluation, none individually included all of the following:
    - Clear ToRs, developed with interested stakeholders, that include a clear expression of the purpose, objectives, expected outputs, uses and users of the evaluation
    - Meeting between evaluators and clients to clarify the ToRs and expected outputs
    - Collation and analysis of relevant literature and secondary information sources that can inform the evaluation
    - Workshop with stakeholders to develop a common understanding of the evaluation, roles and responsibilities and methods to be used, as well as logistics
- Stakeholder analysis
  - Context analysis (local, national, international), including locating the programme within the bigger picture (trade, food security, extension system, climate change...)
  - Construction/re-construction/confirmation of project logic and hypotheses
  - Development of evaluation questions that respond to the logic/hypotheses and to the ToRs
  - Development (with participation of project staff if appropriate) of an appropriate sampling framework that includes a counterfactual
  - Justification for, and description of, the selected set of balanced qualitative and quantitative enquiry tools that provide expression of the perspectives of all pertinent stakeholders
  - Training in the use of concepts and tools
  - Field testing of tools
  - Implementation of data collection
  - Capacity building in the analysis of information collected
  - Description of the analysis of the information collected
  - Draft reporting of the findings (including attribution issues and whether the project is doing things right and doing the right thing)

- Postfieldworkworkshop(s)with stakeholders and project staff to feed back draft findings, confirm their veracity, confirm or modify the viability of the recommendations, and develop an action plan (what, who, when, where, with what resources)
- De-briefing with clients
- Finalisationofreportincludinganexecutive summary (with translation into the local language if appropriate), clear presentation of findings (including tables, figures and photos), justification and description of the approach and rationale for the methods/tools employed, analysis of the limitations of the methods used, project logic, evaluation questions, recommendations for each user, and an agreed action plan for implementation of the recommendations
- Dissemination of the findings using a range of mechanisms/media to reach different audiences
- A relatively narrow set of methods was used by the case studies (despite our efforts to include a wide range). None included outcome mapping, Most Significant Change (although Appreciative Inquiry used in #1 has some similarities), or the Balanced Scorecard Approach;
- Evaluation is enhanced where programmes adopted an evaluation cul-

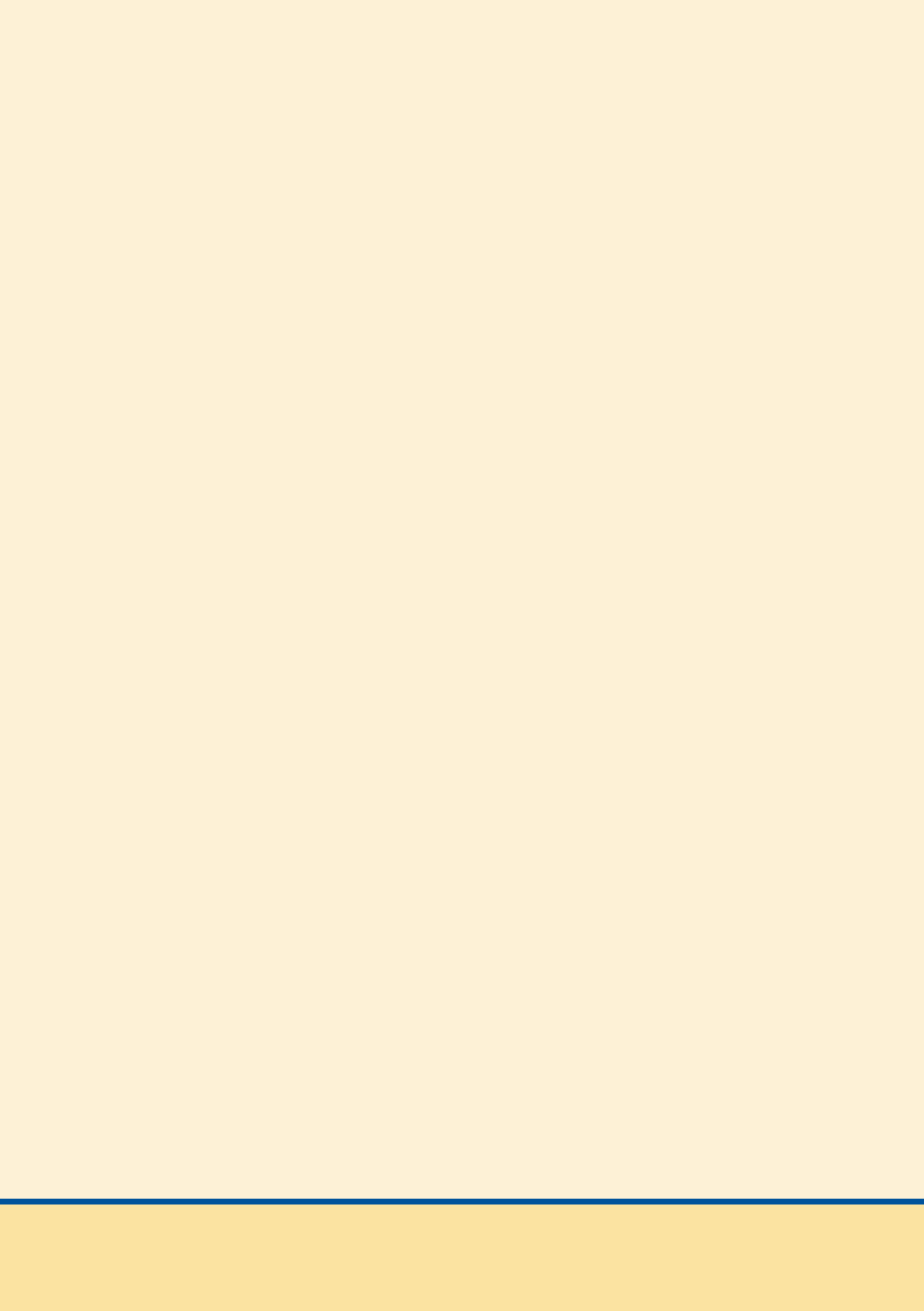
ture from design through implementation and had a results-based framework against which to evaluate progress, change and potential. Project and beneficiary monitoring data, reports on internal reflection or evaluation events, quarterly or six-monthly reporting against milestones and indicators, and a baseline that is related to project indicators also help ensure that the evaluation provides an accurate and comprehensive assessment of the situation and makes appropriate recommendations.

## Use of this Meta-Analysis

It is intended that the information presented here from the 17 case studies will be combined with more theoretical considerations from a meta-review of evaluation literature in the development of an evaluation guidance kit and a policy brief. The guidance kit will also benefit from dialogue held with acknowledged specialists in extension evaluation, from the deliberations of a workshop on extension evaluation to be held in Greenwich, UK in September 2010, and from discussions with extension and evaluation specialists at the First Intercontinental Meeting of the Global Forum for Rural Advisory Services, held in Santiago, Chile in November 2010.

# Annexes







## Annex 1: The Matrix used in the interrogation of the 17 Case studies

	Category	Responses
<b>1.</b>	<b>Validation</b>	
<b>1.1</b>	<b>Problem Definition</b>	
<b>1.1.1</b>	<b>Purpose of the evaluation</b>	
	<ul style="list-style-type: none"> <li>a. Are objectives clear?</li> <li>b. Are users defined?</li> <li>c. Are uses defined?</li> </ul>	
<b>1.1.2</b>	<b>How has the problem been defined and further developed in evaluation questions?</b>	
	<ul style="list-style-type: none"> <li>a. Is the purpose clearly defined?</li> <li>b. Do the evaluation criteria relate to the purpose of the evaluation?</li> <li>c. Is the purpose relatively narrow?</li> <li>d. Have the evaluators raised broader issues?</li> <li>e. Has the evaluator described how the goals of extension are understood?</li> <li>f. What methods were used to define the evaluation criteria and questions?</li> <li>g. With whom were they defined?</li> </ul>	
<b>1.1.3</b>	<b>Nature, clarity and consistency of evaluation criteria</b>	
	<ul style="list-style-type: none"> <li>a. A clear and comprehensive description of the evaluation criteria in terms of the five OECD DAC criteria – relevance, efficiency, effectiveness, impact &amp; sustainability</li> </ul>	

	Category	Responses
<b>1.2</b>	<b>Subject evaluated</b>	
<b>1.2.1</b>	<b>Definition, functionality, and parameters of the subject evaluated</b>	
	<ul style="list-style-type: none"> <li><b>a.</b> A detailed description of the component activities evaluated (type, target group, location, period, organisation, financial value, etc.) – the ‘evaluation population’</li> <li><b>b.</b> Does the above reveal a different set of objectives, or a limited set of actors?</li> <li><b>c.</b> What methods were used to define the evaluation population?</li> </ul>	
<b>1.2.2</b>	<b>The place of the subject evaluated in its policy and institutional context</b>	
	<ul style="list-style-type: none"> <li><b>a.</b> An account of relevant policy contexts and principles and of the institutional environment in which the subject evaluated operates</li> <li><b>b.</b> Does the evaluation ask critical questions about whether the range of actors reflects a commitment to pluralistic policies</li> <li><b>c.</b> Do such policies exist in the country?</li> <li><b>d.</b> Was the programme evaluated based on its contextual relevance or its relevance to the needs of users</li> <li><b>e.</b> What methods were used for contextual analysis</li> </ul>	

	Category	Responses
<b>1.3</b>	<b>Policy theory</b>	
<b>1.3.1</b>	<b>Account or reconstruction of intervention logic and result levels</b>	
	<ul style="list-style-type: none"> <li><b>a.</b> An account of the theory behind the policy, including assumptions about causal and final relationships underpinning the activities evaluated, and assumptions about the input/output/outcome/impact hierarchy</li> <li><b>b.</b> Does this provide a basis for comparing the programme's intervention logic with broader policy norms that have (ostensibly) been adopted by the programme partners?</li> <li><b>c.</b> Does it allow assessment of coherence or inconsistency between a policy position and programme design and implementation?</li> <li><b>d.</b> What methods were used to reconstruct and interrogate the intervention logic</li> </ul>	
<b>1.3.2</b>	<b>Operationalisation of results measurement via indicators</b>	
	<ul style="list-style-type: none"> <li><b>a.</b> a) To what extent have the indicators defined at the various result levels ultimately measured the aspects of the programme that are relevant?</li> <li><b>a.</b> b) Is the 'theory of change' of the programme reflected in the indicators?</li> <li><b>a.</b> c) Do the indicators steer implementation</li> </ul>	

	Category	Responses
<b>1.4</b>	<b>Analysis</b>	
<b>1.4.1</b>	<b>Attribution</b>	
	<ul style="list-style-type: none"> <li><b>a.</b> How have the evaluators dealt with the inherent challenges in attributing wider outcomes and impacts to interventions?</li> <li><b>b.</b> To what extent have the evaluators recognised these challenges and found ways to address them?</li> <li><b>c.</b> How clearly have the intervention activities been defined?</li> <li><b>d.</b> By what methods do they assess whether the intervention activities used were effective in bringing about the desired changes</li> <li><b>e.</b> By what methods do they assess whether these have translated into impacts e.g. increased income, poverty alleviation...</li> <li><b>f.</b> How do they extrapolate impacts into the future</li> </ul>	
<b>1.4.2</b>	<b>Information sources, information collection, and information processing</b>	
	<ul style="list-style-type: none"> <li><b>a.</b> Was the information needed available to the evaluators?</li> <li><b>b.</b> By what methods was missing data collected?</li> </ul>	
<b>1.4.3</b>	<b>Underpinning of conclusions by results.</b>	
	<ul style="list-style-type: none"> <li><b>a.</b> a) To what extent do the conclusions arise from the evaluation results/ findings?</li> </ul>	
<b>2.</b>	<b>Reliability (how reliable are the evaluation results)</b>	
<b>2.1.</b>	<b>Evaluation methods</b>	
<b>2.1.1</b>	<b>Specification of and justification for evaluation methods applied</b>	
	<ul style="list-style-type: none"> <li><b>a.</b> A precise description of and justification for the evaluation methods and techniques applied</li> <li><b>b.</b> If this is absent what can be concluded about the choice of methods – e.g. the sampling framework and its suitability?</li> <li><b>c.</b> What is the relative weight given to qualitative and quantitative methods?</li> </ul>	

	Category	Responses
<b>2.1.2</b>	<b>Verification of information / triangulation</b>	
	<ul style="list-style-type: none"> <li>a. How is information checked, used, and applied to collect information about the same features and phenomena</li> <li>b. Is there a counterfactual?</li> <li>c. What factors influence the choice of “experimental method” against participatory action research approaches?</li> <li>d. What methods are used to construct comparisons, especially if baselines are missing?</li> </ul>	
<b>2.2</b>	<b>Scope</b>	
<b>2.2.1</b>	<b>Representativeness of sample or case study selection</b>	
	<ul style="list-style-type: none"> <li>a. The extent to which the conclusions drawn from the sample evaluated or case study conducted apply to the entire ‘evaluation population’</li> <li>b. Are there biases toward samples that are easily identified as ‘extension’ (e.g. the public sector) or easily measured, at the expense of important aspects of the extension system that are difficult to analyse?</li> <li>c. What methods &amp; tools were used in the evaluation to examine broader and informal aspects such as social networks, private sector actors etc; to explore differential outcomes for different social groups - in relation to poverty, gender, age, ethnicity etc</li> </ul>	
<b>2.2.2</b>	<b>Limitations of the evaluation</b>	
	<ul style="list-style-type: none"> <li>a. What were the shortcomings in the evaluation and limitations regarding the extent to which the results and conclusions can be generalised?</li> </ul>	

	Category	Responses
<b>3.</b>	<b>Usability (of evaluation results)</b>	
<b>3.1</b>	<b>Presentation</b>	
<b>3.1.2</b>	<b>Accessibility of the evaluation results</b>	
	<ul style="list-style-type: none"> <li><b>a.</b> How clearly and completely does the evaluation report and its summary reflect the essence of the evaluation, especially its main results</li> <li><b>b.</b> Is the report appropriately written for its various audiences</li> <li><b>c.</b> What methods or media have been usefully employed to convey the results?</li> </ul>	
<b>3.2</b>	<b>Connections (logic)</b>	
<b>3.2.1</b>	<b>Evaluation questions answered by conclusions</b>	
	<ul style="list-style-type: none"> <li><b>a.</b> Are all the evaluation questions answered by the conclusions?</li> <li><b>b.</b> If not, is this a problem of the methods and tools used?</li> </ul>	
<b>3.2.2</b>	<b>Feasibility/relevance of lessons or recommendations</b>	
	<ul style="list-style-type: none"> <li><b>a.</b> Are the recommendations feasible?</li> <li><b>b.</b> Do they lie within the remit of those responsible to act, especially the entity that commissioned the evaluation?</li> <li><b>c.</b> Do the recommendations support greater pluralism?</li> </ul>	
<b>3.3</b>	<b>Partnership</b>	
<b>3.3.1</b>	<b>Capacity development</b>	
	<ul style="list-style-type: none"> <li><b>a.</b> How has the evaluation contributed to local capacities for future monitoring and evaluation?</li> <li><b>b.</b> Were local actors involved in the evaluation?</li> <li><b>c.</b> As a result, do local actors now have the capacity to conduct future evaluations?</li> <li><b>d.</b> What methods were used to promote partnership and capacity building?</li> </ul>	

	Category	Responses
<b>3.3.2</b>	<b>Local ownership and engagement</b>	
	<ul style="list-style-type: none"> <li>a. How was the evaluation used as a catalyst for mobilising local interest in continued critical assessment of achievements</li> <li>b. Was the evaluation designed to link local actors interested in reforming extension policy and systems?</li> <li>c. Does the evaluation analyse local ownership and financial sustainability?</li> <li>d. What methods were used to investigate social, environmental, economic and institutional sustainability?</li> </ul>	
<b>3.3.3</b>	<b>Coordination and alignment</b>	
	<ul style="list-style-type: none"> <li>a. How was the evaluation linked to broader, nationally-owned processes of assessing the effectiveness of aid and relevant programming areas?</li> <li>b. Was the evaluation structured so as to optimally contribute to national processes, especially in relevant areas that are not always associated with agriculture per se, e.g., climate change, food security, nutrition, gender equality, market development?</li> </ul>	
<b>3.3.4</b>	<b>Governance and management</b>	
	<ul style="list-style-type: none"> <li>a. To what extent did the evaluation have a governance/management structure that encouraged local ownership while maintaining the evaluators' independence?</li> </ul>	
<b>3.3.5</b>	<b>Incorporation of stakeholder perspectives and comments</b>	
	<ul style="list-style-type: none"> <li>a. How has the evaluation brought in the perspective of local stakeholders in the report as well as including their reactions to the draft report?</li> <li>b. What feedback on the evaluation was given to those evaluated and to other stakeholders to confirm/refute/modify draft findings?</li> </ul>	

## Annex 2: Methods used in each case study evaluation

#	Project / Programme	Methods employed
1	Agricultural Support Programme, Zambia	<ul style="list-style-type: none"> <li>● <b>Appreciative Inquiry</b> coupled to a set of ten main evaluation questions (“perspectives”).</li> <li>● <b>Ranking:</b> Different extension interventions are ranked for their effectiveness in bringing about positive change</li> </ul>
2	National Agricultural Advisory Services (NAADS), Uganda	<ul style="list-style-type: none"> <li>● <b>Questionnaire:</b> Internal NAADS District survey in 4 Districts surveyed 1555 households and 305 NAADS farmer groups</li> <li>● <b>Questionnaire:</b> External survey covering 141 NAADS farmer groups and 54 non-NAADS farmer groups (counterfactual).</li> <li>● <b>Questionnaire:</b> UoG Bureau of Statistics crop, service delivery and household surveys</li> <li>● <b>Interviews and observation:</b> Visits by MPs to 2 Districts to gauge the effectiveness of NAADS and the need for policy support.</li> <li>● <b>Stakeholder workshop:</b> To present and discuss findings of the above.</li> </ul>
3	OXFAM Fair trade, multi-region	<ul style="list-style-type: none"> <li>● Economic Impact Ratio: Compares the income from Fair Trade with the opportunity cost of labour (the income they would have got without FT, which represents the “without project” scenario).</li> <li>● Videos made for wider dissemination of results</li> </ul>
4	Ten years work of Misereor partners, Uganda	<ul style="list-style-type: none"> <li>● Pre-fieldwork workshop helped define questionnaire questions</li> <li>● Questionnaire: 3 sample groups: early and late adopters within project and non-project. 20% sample; randomly selected</li> <li>● Post-survey workshop for feedback and exploration of future scenarios</li> </ul>



Comments
<ul style="list-style-type: none"> <li>• Qualitative only</li> <li>• Participatory</li> <li>• No sampling framework</li> <li>• Interacted with wide range of stakeholders</li> <li>• No counterfactual or any other basis for comparison</li> <li>• No baseline- No quantitative information</li> <li>• Evaluation objectives explicitly include stakeholder learning</li> <li>• Feedback to stakeholders (information, confirmation) and clients</li> </ul>
<ul style="list-style-type: none"> <li>• External survey by Scanagri used a counterfactual</li> <li>• Several sets of evaluation criteria (corresponding to the different components). Some cohesion or nesting b/w them would have been useful</li> <li>• No description of the actual methods used or any sampling framework used</li> <li>• NAADS has a logical framework against which to measure progress</li> <li>• Evaluation is against baseline data and targets, but also mindful of changes to the implementation schedule due to political pressures</li> </ul>
<ul style="list-style-type: none"> <li>• No baseline information, but producer groups kept records of FT income, which allowed calculation of the Economic Impact Ratio</li> <li>• Looked across the value chain, not just at the producers, because the effectiveness of the value chain influences the impact on producers</li> <li>• No feedback workshop</li> <li>• No sampling framework mentioned</li> </ul>
<ul style="list-style-type: none"> <li>• Sampling framework</li> <li>• Good counterfactual</li> <li>• Good process for field testing the questionnaire</li> <li>• Narrow range of stakeholders surveyed</li> </ul>

#	Project / Programme	Methods employed
5	Rural Advisory Service and Kyrgyz Swiss Agricultural Program	<ul style="list-style-type: none"> <li>• Review of Program Document and any other relevant documents</li> <li>• Interviews with representatives of SDC, WB/IFAD, KSAP, RAS, MAWRPI, and stakeholders and beneficiaries of the various KSAP components.</li> <li>• Field visits to selected districts where RAS works</li> <li>• Presentation of results to local stakeholders and debriefing to donors</li> </ul>
6	Southern Highlands Dairy Development Project, Tanzania	<ul style="list-style-type: none"> <li>• Review of project literature and other relevant documents</li> <li>• Defined hypotheses and then questions to prove or disprove hypotheses</li> <li>• Interviews (57–0.5% sample of farmers + government, NGO and project staff)</li> <li>• Feedback workshop with wide representation of stakeholders (including farmers)</li> </ul>
7	Agricultural Extension Services in Indonesia	<ul style="list-style-type: none"> <li>• Indicators (39) corresponding to the project expected outputs.</li> <li>• Questionnaires used for a Benchmark survey (Baseline) and an End-of-Project (EOP) survey: Three samples were used: 450 DAFEP households (from a population of 30,000); 450 non-DAFEP households; spillover households within the DAFEP Districts.</li> <li>• Villages selected at random and then households at random within these.</li> </ul>
8	Public Agricultural Extension System of Ethiopia	<ul style="list-style-type: none"> <li>• Desk review of relevant literature</li> <li>• Key informant interviews</li> <li>• Stakeholder consultations using focus group discussions: Interviews with farmers and farmer groups, regional heads, office heads, subject matter specialists (SMSS), and DAs for local extension experiences. At the ATVETs, administrators, instructors, and students were interviewed</li> <li>• Field visits: six regions and nine woredas were identified by criteria that would allow the study to cover a diverse set of agro-ecologies, regions, and production systems with the time and personnel available, and to reflect a wide range of local extension and ATVET experiences.</li> <li>• Post survey workshops with stakeholders</li> </ul>

Comments	
	<ul style="list-style-type: none"> <li>• Previous External Review acts as a baseline for comparison</li> <li>• Actors interviewed mostly managers and funders, rather than beneficiaries/end users of extension</li> <li>• No methodology section (so no sampling framework or interview methods)</li> <li>• No counterfactual</li> </ul>
	<ul style="list-style-type: none"> <li>• Narrow stakeholders (ignores private sector and national level staff)</li> <li>• Little useful quantitative baseline data despite the project age (25 yrs)</li> <li>• No sampling framework</li> </ul>
	<ul style="list-style-type: none"> <li>• High level of rigour</li> <li>• Sophisticated data analysis</li> <li>• Good counterfactual</li> <li>• Poor articulation of objectives, uses, users and purpose</li> <li>• No feedback workshop for local stakeholders</li> </ul>
	<ul style="list-style-type: none"> <li>• A pre-test of data collection instruments was done</li> <li>• Continual briefings of and feedback from the MOARD, a panel of Ethiopian development experts, and other stakeholders</li> <li>• No sampling framework mentioned although there were criteria for the selection of woredas and groups</li> </ul>

#	Project / Programme	Methods employed
9	Farmer Field Schools in East Africa	<ul style="list-style-type: none"> <li>• Document review</li> <li>• Semi-structured interviews with key informants.</li> <li>• A longitudinal impact evaluation with quasi-experimental methods (propensity score matching and covariate matching).</li> <li>• 1,126 households were randomly selected from villages with FFSs and villages without FFSs. The villages without FFSs chosen were comparable in terms of biophysical and socioeconomic characteristics</li> <li>• Twenty (20) FFSs per country were randomly selected from purposively selected districts.</li> <li>• Farmers selected were proportional to the FFSs in each district and diversity of agro-ecological zones.</li> <li>• Household members were randomly selected, with those interviewed proportional to the total membership in FFSs. Non-FFS participants were randomly sampled.</li> <li>• Closed-ended questions were used and analysis of responses done using probit regression model.</li> <li>• Data were checked using data-cleaning syntax that checked for errors.</li> </ul>
10	Extension Services and Market Regulation in the EU	<ul style="list-style-type: none"> <li>• Review of available documents: monographs, comparative studies, reports, guidelines, management plans.</li> <li>• Interviews in 2000 and 2002 with main institutions like chambers of agriculture, agricultural unions, national agencies, technical institutes, rural development bodies, etc. based on a list of open questions on ten topics</li> <li>• An institutional analysis was done by comparing the evolution of institutional configurations according to the systems for which the technical support were structured.</li> <li>• An organizational analysis of how the institutions surveyed understand current trends and adjust their internal activities and functions accordingly.</li> </ul>

## Comments

- Qualitative and quantitative methods were used to collect data.
- Used a counterfactual: The impact was estimated by subtracting the changes in yield of FFS nonparticipants before and after FFS from the change in yield of the FFS participants

- Uses and users not specified
- No sampling framework given
- There is no evidence of project log frames, indicators, targets or milestones in the evaluation
- No feedback workshop

#	Project / Programme	Methods employed
11	Plant Clinics in Bangladesh	<ul style="list-style-type: none"> <li>• Questionnaire: 350 farmers from all 18 clinics were interviewed using a 1-page questionnaire and were selected opportunistically.</li> <li>• The survey team asked the farmers what the clinic had recommended, what they did after receiving it, compared their answers with the clinic register and classified each farmer according to how well he remembered the clinic recommendations. The team then decided if the farmer had adopted the recommendation or not.</li> <li>• Data collected were entered in MS Excel spreadsheet and analyzed using MS Excel and SPSS by consultants from AAS.</li> <li>• Descriptive statistics, mean, proportion and Students T Test were performed to compare before and after adoption of plant health services.</li> </ul>
12	Laos Extension for Agriculture Project	<ul style="list-style-type: none"> <li>• Review of previous reports</li> <li>• Interviews with key informants</li> <li>• Follow-up discussions with LEAP project staff</li> <li>• De-briefing sessions with SDC and NAFES (funding and government agencies)</li> </ul>
13	The Nat Agric and Livestock Extension Programme, Kenya	<ul style="list-style-type: none"> <li>• Review of documentation (including audit reports, internal budgets, District and Division agriculture statistics, and internally commissioned monitoring and evaluation documents)</li> <li>• Discussions and interviews with SIDA staff in Stockholm and Nairobi</li> <li>• Interviews of male and female farmers and leaders at different levels and other stakeholders in agricultural extension</li> <li>• Focus group discussions, including SWOT analysis</li> <li>• Participatory force-field analysis focusing on farmers' relations to various institutions and organisations</li> <li>• Workshop to discuss the draft report with major stakeholders (about 50 participants)</li> </ul>

## Comments

- Questionnaire was field tested before application
- No counterfactual, but impact on farmer production cost, yield and income before adoption of clinic recommendations and after was assessed.
- No feedback workshop

- Clear and simple objectives, users and uses.
- 5 simple evaluation questions
- Clear use of DAC criteria
- Insufficient time for rigorous analysis of activities
- No clear sampling framework
- No logframe or indicators
- No quantitative information
- No baseline against which to measure change

- No quantitative information
- No clear sampling framework
- Good treatment of 5 DAC criteria

#	Project / Programme	Methods employed
14	Farm Women in Development, India	<ul style="list-style-type: none"> <li>• Review of documents (project documents, evaluations and reviews)</li> <li>• Analysis of project data (costs, achievements)</li> <li>• Stakeholder interviews (with Danida staff, national and state government officials and project staff).</li> <li>• Week-long planning workshop, prior to the fieldwork</li> <li>• Questionnaires which reflected both the time dimension (different phases) and geographical coverage.</li> <li>• Key informant interviews supplemented the questionnaires</li> <li>• Findings focussed on Knowledge, Attitude and Practice (KAP) changes, cost effectiveness, economic, social, institutional and policy changes and sustainability</li> </ul>
15	Videos that Strengthen Rural Women's Capability to Innovate, Bangladesh	<ul style="list-style-type: none"> <li>• Knowledge, Attitude and Practice baseline survey and post post-intervention survey (six months later) carried out in 12 districts of Bangladesh, in 50 villages with 1,252 women.</li> <li>• 7 villages were used as control.</li> <li>• During their visits to the villages, the teams also collected qualitative data. They probed for and documented additional local innovations and asked people for feedback on the videos.</li> </ul>
16	Farmer Trainer Programme in Meru, Kenya	<ul style="list-style-type: none"> <li>• Cost: benefit evaluation (financial and non-financial costs and benefits – including non-tangibles like prestige and knowledge; also direct and secondary costs and benefits)</li> <li>• Interviews with stakeholders</li> <li>• Gross Margin Analysis for income/costs of trainees</li> <li>• Internal Rate of Return for assessing rate of return to project investment (270%)</li> <li>• Data on costs incurred by farmer trainers were collected from Monitoring records of fifteen randomly selected farmer trainers.</li> <li>• Data on the costs and revenues of dairy goat production were also available from the project.</li> </ul>



## Comments

- No baseline (even though project 15-20 years old).
- No description of the ways in which the women, groups and organisations consulted were representative (sampling framework)
- No counterfactual
- No logframe or indicators
- No mechanism allowing reaction (feedback) to draft report
- Focus on gender rather than on women
- Evaluates against all 5 DAC criteria
- Looks across different scales: local, state and national

- Counterfactual
- No feedback events with local stakeholders or commissioning organisation
- Good link to TV (showing of videos)

*"In 2005, under the GSI project, we handed over a video compact disc (VCD) after each show to a person selected by the women in the audience. Interestingly, the women mostly identified a man to look after the CD. Within a very short time, and at no cost to the project, they had organized about 140 more shows for other farmers in their villages. Overall, each CD distributed has triggered changes in the knowledge and practices of about 200 farmers without too much effort."*

- No social or wealth differentiation of benefits
- No feedback workshop

#	Project / Programme	Methods employed
17	Fodder trees on milk production and income in East Africa	<ul style="list-style-type: none"> <li>• Review of documents: previous studies and newly generated and analyzed empirical data that has not previously been published</li> <li>• Community workshops (to share the study objectives with the farmers, the type of information sought and the methods to be used for data gathering)</li> <li>• Participatory rural appraisal (PRA)</li> <li>• Focus group discussions</li> <li>• Case studies</li> <li>• Interviews with key informants</li> <li>• The farmers were selected by stratified random sampling.</li> <li>• 15 farmers in every division who were not adopters were randomly picked by the enumerators</li> </ul>

## Comments

- The data were collected from a single recall, with attendant limitations
- Participatory, with good involvement of a range of local stakeholders
- Counterfactual
- No feedback mechanism

GFRAS – Global Forum for Rural Advisory Services  
c/o Agridea  
Eschikon 28  
8315 Lindau  
Switzerland  
Tel: +41 (0)52 354 97 64



Except where otherwise noted, this work is licensed under  
<http://creativecommons.org/licenses/by-nc/3.0/>

The Global Forum for Rural Advisory Services (GFRAS) has commissioned the Natural Resources Institute to develop a toolkit for the evaluation of extension (projects, programmes, tools and initiatives). This commission has a number of components:

A meta-evaluation of 15-20 evaluation case studies (presented here)

- A meta-review of the literature relevant to extension evaluation methods
- A workshop with practitioners and experienced evaluators to discuss the findings of a) and b) and to identify an initial set of tools
- A proposal for testing the proposed tools in a second phase of the project
- A brief of the toolkit for policymakers.

The overall purpose of this project is to identify methods for better evaluation of extension through the development of a toolkit for extension evaluation. The meta-evaluation and meta-review will also provide an in-depth basis for the selection of the approaches, methods and tools in the toolkit.