global forum for rural advisory services
forum mondial pour le conseil rural
foro global para los servicios de asesoría rural

Natural Resources Institute

Review of Literature on Evaluation Methods Relevant to Extension

Adrienne Martin, Sabine Gündel, Essie Apenteng, and Barry Pound

2011

www.g-fras.org

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
Adrienne Martin*
Sabine Gündel
Essie Apenteng
Barry Pound

Design & Layout:
Ana María Hintermann-Villamil,
webhint.ch Gümligen, Switzerland

2011, Lindau, Switzerland

** Lead author: Natural Resources Institute, University of Greenwich, Chatham, Kent ME44TB, UK. A.M.Martin@gre.ac.uk*

Introduction

The Global Forum for Rural Advisory Services (GFRAS) has commissioned work to develop guidance for the evaluation of extension (projects, programmes, tools and initiatives). The purpose of the overall project is to identify methods for better evaluation of extension through the development of a Guidance Kit for extension evaluation.

This review of literature on evaluation methods, in combination with a meta-evaluation[1] of extension evaluation[2] case studies , is intended to be a resource for developing the Guidance Kit. It is envisaged that this paper will be of interest to those involved in managing and implementing evaluations of rural advisory services as well as to extension and evaluation specialists.

The literature review focuses specifically on approaches and methodologies in evaluation which are relevant for evaluating initiatives in extension or rural advisory services. The review adopts the following definition of extension:

Rural advisory services, also called extension, are 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 (GFRAS 2010a).

This includes the processes and actions which facilitate access and interaction with and among partners in research, education, agribusiness and other relevant institutions (Christoplos 2010). It also includes capacity-strengthening functions and the interface between extension and agricultural research, such as action learning approaches and participatory research. However, the review does not specifically address the evaluation of other rural services such as health and education.

The definition of evaluation, from the OECD Development Assistance Committee is the basis for the discussion of evaluation, although as will be explained, there are different types of evaluation.

An evaluation is an assessment, as systematic and objective as possible, of an ongoing or completed project, programme or policy, its design, implementation and results. The aim is to determine the relevance and fulfilment of objectives, devel-

[1]Meta-evaluation is an evaluation of an evaluation or a number of evaluations. It examines their utility, feasibility, propriety and accuracy in order to guide the evaluation and to publicly report its strengths and weaknesses (Stufflebeam 2000).
[2]Pound, B., S. Gündel, A. Martin, and E. Apenteng. 2011. Meta-evaluation of extension evaluation case studies, Lindau: NRI/GFRAS.

opmental efficiency, effectiveness, impact and sustainability. An evaluation should provide information that is credible and useful, enabling the incorporation of lessons learned into the decision-making process of both recipients and donors (OECD 1991, 5)

The materials reviewed include academic papers, existing evaluation toolboxes and major evaluation guidelines and frameworks. Firstly, the context and scope of the review are discussed, followed by sections addressing the purposes, users and uses of evaluation, evaluation standards and criteria, approaches, rigour and attribution. The final three sections discuss the principles for evaluation of rural advisory services in highly complex situations, the strengths, weaknesses and gaps in existing approaches, and the ways forward.

The changing orientation of extension

Extension services in the past were closely linked to the agricultural sector and to agricultural development programmes. Despite signing up for the United Nations 'halving hunger' goals in 1996 and again in 2000, donor countries reduced official aid to agriculture from 16.8% of all official development spending in 1979, to just 3.4% in 2004 (ActionAid 2009). In addition, this trend was mirrored by reductions in developing country national budget allocations for agriculture (Global Donor Platform for Rural Development 2008). Agriculture has suffered drastic declines in development investment and assistance (Global Donor Platform for Rural Development 2006), with far-reaching consequences for agricultural infrastructure, including extension services. Over the last decade, projects and programmes in agricultural extension have been to a large extent superseded by other development approaches such as environmental protection, good governance, poverty reduction, etc. (World Bank 2008). These initiatives were often established as separate programmes rather than building on and integrating with existing rural services.

These trends have resulted in severe under-funding or closure of national extension services and also in a lack of research and publications in this field. There is now a renewed interest in, and recognition of, the role of agriculture in food security and wider environmental services, which has put this sector back on the global agenda (Nagel 2003; Anderson 2007; World Bank 2008). Whereas in the past the focus was on production increases through modern agricultural technologies, the recent focus takes into account the wider set of functions agriculture has in provision of ecosystem services and in rural livelihoods.

Birner et al. (2006) describe the shift in focus as follows: *"This shift emphasizes the continued need for agricultural advisory services as a means of promoting agricultural productivity, increasing food security, improving rural livelihoods, and promoting agriculture as an engine of propoor economic growth. Agricultural advisory services are also needed to meet the new challenges agriculture is confronted with: changes in the global food and agricultural system, including the rise of supermarkets and the growing importance of standards and labels; growth in non-farm rural employment and agribusiness; constraints imposed by HIV/AIDS, and other health challenges that affect rural livelihoods; and the deterioration of the natural resource base and climate change."*

This shift in focus is also reflected in the changing perspective on extension systems. Whereas production-oriented agricultural focus was served by linear, commodity-focused, top-down extension approaches (e.g. 1970s and 1980s, the Training and Visit system (T&V)) and later, by more participatory bottom-up approaches such as Farmer-Field Schools (FFSs), the recent focus on the multi-functionality of agriculture requires a more pluralistic and demand-driven extension or rural advisory approach, *"accounting for the complexity of agricultural systems within diverse social and ecological contexts"* (IAASTD 2009, 4). This diversity of contexts requires regional or even local extension approaches, as no single approach can simultaneously increase market orientation, food security and mitigate climate change (Christoplos 2010).

The definition below by Leeuwis and van den Ban (2004) emphasises the complexity in terms of the actors and innovation requirements involved.

Extension [is] a series of embedded communicative interventions that are meant, among others, to develop and/or induce innovations which supposedly help to resolve (usually multi-actor) problematic situations (Leeuwis and van den Ban 2004).

In addition to the above, The Common Framework on Agricultural Extension (Neuchâtel Group 1999 and 2000) emphasises the following drivers for change in extension services:

- Decentralization, liberalization, privatisation, and democratization
- New actors are becoming involved in "extension" activities
- Public spending on extension is shrinking
- The aims of official development assistance are becoming more focused

GFRAS interprets Rural Advisory Services (RAS) as follows:

RAS are about strengthening capacities, empowering rural people, and promoting innovations. RAS support people to obtain skills and information, and to address challenges so as to improve their livelihoods and well-being. Traditionally, RAS disseminate information about technologies, markets, inputs and financial services, and assist farmers to develop their farming and management skills. But RAS also broker interactions between farmers, the private sector, research, education, and government. RAS coach different actors to improve market access, deal with changing patterns of risk, and protect the environment.'

Source: GFRAS web site, http://www.g-fras.org/

Furthermore, the Neuchâtel Group (2006) identified the following principles for Demand-Driven Service Delivery Systems:

- Services shall be driven by user demand
- Service providers shall be accountable to the users
- Users shall have a free choice of service providers

The change in terminology tries to capture the more holistic perspective and complexities of modern extension service, although so far it seems difficult to leave the well-established terminology behind.

The changing orientation of evaluation

The importance of evaluation has long been recognized, but there have been some important shifts in the understanding of its function and significance in the context of globalization and changing policy objectives, international aid modalities, and the Paris Declaration and Accra Agenda for Action, with their focus on delivering and accounting for results (OECD 2005/2008). Evaluation practice has had to adjust to accommodate this level of complexity. It has to play its traditional role of generating information on the implementation and results of a programme or project, but in addition assess policy impacts and also provide the basis for improved management and decision-making as well as for transparent accountability to citizens (Commission of the European Communities 2007).

There have been shifts in international donor support to agriculture, from project-based approaches to sector support programmes (e.g. Uganda, Rwanda, Tanzania and Mozambique), trends towards decentralisation of agricultural service provision through local government, and a new emphasis on the inter-connectedness of local agricultural development

with wider national and international processes. A further change has been in the conceptualization of extension as involving complex processes of innovation – multiple and new sources of information, multiple and differentiated actors, market-driven trends in production and supply, linked to national and international trade policy and regulations. Similarly, the recognition of the range of actors participating has widened to include private sector companies, non-governmental bodies, national and international NGOs and charitable foundations, and a range of government service providers and political actors. Accompanying this has been the mainstreaming of concerns for capacity strengthening among actors at different levels, and for the 'empowerment' of farmers, particularly through farmers' groups and associations.

The changes have required a different approach to evaluation of agricultural development projects, including those involved in agricultural extension and advisory service provision. These have included shifts from single technology focused evaluations, e.g. introduction of a new crop variety or new methods of crop or livestock management, to more complex assessments of changes in agricultural institutions, farming systems, human resource capacity and trade relationships, together with the impact on the incomes and livelihoods of participants. Evaluations are also concerned with the effects of policy changes relating to agricultural service provision and the extent to which these foster flexible and equitable relationships among different service providers, including the private sector (Swanson and Rajalahti (2010). A growing interest in participatory evaluation (Rajalahti et al. 2005, 35) has developed, partly in response to the empowerment agenda.

It is also clearly recognized that there is a difference between evaluations which are intended to account to investors and the public (including donors, governments, tax payers, farmers' organizations, etc.) for money that has been appropriately spent in accordance with agreed deliverables providing a satisfactory return on investment, and those which are more learning-oriented and concerned with what works and what doesn't work and what adjustments should be made.

There have been different phases of donor interest in evaluation theory and approaches. The current resurgence of interest in evaluation and impact assessment in general, is linked in part to the international financial crisis, the need for expenditure cuts and a renewed emphasis on value for money. It is consistent with the emphasis on measuring progress and development effectiveness articulated in the Paris declaration (OECD/DAC Working Party on Aid Effectiveness 2008).

The scope of evaluation practice has extended to more sophisticated ex ante evaluation of proposed projects, more emphasis on outcomes and impacts rather than the immediate outputs of projects – particularly in view of concerns on aid effectiveness and contribution to the MDGs. There have been considerable advances in the theory and practice of impact evaluation and a particular interest in evaluations which can attribute changes in selected outcomes to a specific intervention. This reflects an interest in understanding whether development interventions actually have a significant effect on outcomes and whether they are cost effective.

Thus there has been a co-existence of methods and approaches in evaluation – some providing detailed quantitative assessments of economic benefits, others using qualitative methods (Patton 2002a) to examine transformation in agricultural extension processes and access to services and to learn lessons about effective approaches to building capacity and empowerment, etc. Underlying these are different values and objectives in evaluation and in development (Alkin and Christie 2004). While it could be argued that the range of methods and approaches provides a more effective choice depending on the context and purpose of an evaluation, in some cases the result has been polarization and contestation concerning evaluation methods, or merely confusion (White 2009a). In particular, methods for assessing projects emphasizing processes of change, complex interrelationships and learning are less well developed and are perceived as lacking the rigour of methods used for assessing the outcomes and impacts of technical interventions. However, it is increasingly argued that multiple and complementary evaluation approaches are needed to address the complexity of RAS.

A recent statement by Anderson (2007) indicates that 'hard' data on newly emerging RAS are still in short supply and are urgently needed to support the functioning of RAS and inform extension policy. He cites the framework proposed by the Extension Research Group at IFPRI for addressing this knowledge gap by assisting the design and analysis of pluralistic extension systems. The aim of this framework is to identify and learn from "best-fit" solutions.

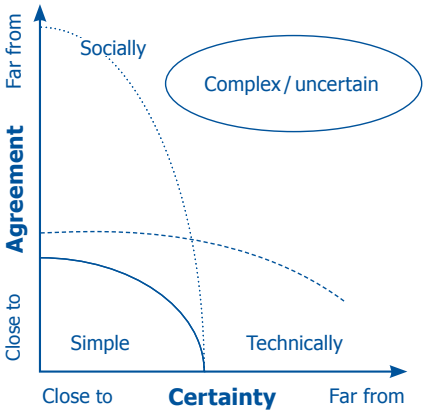
The framework "disentangles" the major characteristics of agricultural advisory services: (1) governance structures, (2) capacity, management and organization, and (3) advisory methods. Four sets of frame conditions have been identified that need to be considered when deciding on these characteristics: (1) the policy environment, (2) the capacity of potential service providers, (3) the type of farming systems and the market access of farm households; and (4) the nature of the local

communities, including their ability to co-operate. The framework suggests an impact-chain approach to analyze the performance and impact of advisory services. The framework can be applied in a dynamic perspective to analyze processes of change over time. Focusing on the question "What works, where, and why?", the framework aims to support a shift from a "best practice" or "one-size-fits-all" to a "best fit" approach in the reform of public advisory services (Anderson 2007; Birner et al. 2006).

Scope of review

Taking as a guiding principle the purpose of GFRAS (2010b), which is *'to provide a space for advocacy and leadership on pluralistic, demand-driven rural advisory services within the global development agenda that promote sustainable rural growth and help the poor',* this review focuses specifically on recent and current literature on evaluation which is relevant to rural advisory services characterized by a high level of social and technical complexity, as noted in Section 2 above.

Figure 1



Source: Patton 2007

The diagram below taken from Patton (2007) depicts a range of extension and/or evaluation situations, which can range from "simple", in terms of their level of social and technical complexity, to "complicated". The social dimension is shown on the vertical axis and technical on the horizontal. As social complexity increases, agreements between the range of different stakeholders are more of a challenge, and with greater technical complexity, technical interventions may require further experimentation and adaptation.

This spectrum could be further increased by taking into account uncertainties such as potential climate tipping-points or global economic developments, which leads to a "complex" situation in which extension needs to function. Such complexity is characterized by dynamic and unpredictable change.

Funnell and Rogers (2011) describe key elements of these different situations.

Table 1 Key elements of simple, complicated and complex situations

	Simple	Complicated	Complex
1. Focus	Single set of objectives	Different objectives valued by different stakeholders: Multiple, competing Objectives at multiple levels of system.	Emerging objectives
2. Governance	Single organization	Specific organizations with formalized requirements	Emergent organizations in flexible ways
3. Consistency	Standardized	Adapted	Adaptive
4. Necessity	Only way to achieve the intended impact	One of several ways to achieve the intended impact	
5. Sufficiency	Sufficient to achieve the intended impact. Works for everyone in the same way.	Only works in conjunction with other interventions (previously, concurrently, or subsequently); Only works in favourable implementation environments	
6. Change trajectory	Simple relationship – readily understood	Complicated relationship– needs expertise to understand and predict	Complex relationship (including tipping points)– cannot be predicted but only understood in retrospect

A range of extension approaches and methods have been developed, implemented and evaluated for "simple" situations, where a specific technology has been promoted by a single agent to a specific target group. Although these situations and approaches may occupy a certain niche in emerging rural advisory services, the main challenges will be encountered in socially complicated situations, with a range of RAS actors and communication technologies and in a context of uncertainty in terms of technological appropriateness and market availability, which results in highly complex situations. For this reason the review concentrates on literature covering the latter situations.

In order to learn from a diverse range of situations we consider in our literature review lessons learned from developed and developing countries as well as sources relating to different sectors. The main sources used in the meta-review are academic papers, existing evaluation toolboxes and major evaluation guidelines and frameworks (e.g. DAC).

The purposes, users and uses of evaluation

Purposes of evaluation

Reflection on objectives, users and uses of evaluation should be the first important step in designing an evaluation approach for extension (Deshler 1997). Conventionally, evaluations were conducted at the end of an intervention (ex post) in order to generate evidence to support claims about its overall achievements (results) which were mainly directed towards donors and policy makers. These evaluations were usually conducted by external evaluators. More recently the purpose of evaluation in extension systems has shifted towards a stronger client/ user focus with a broader set of evaluation objectives, including learning, transparency and capacity development. This shift from a focus purely on *results* to include an evaluation of *process* (in the sense of understanding what is happening in the course and context of a programme or *intervention*), has also tended to increase the involvement of internal evaluators (OECD 2010a).

Different types of evaluation are associated with different objectives and also tend to be associated (though not exclusively) with different phases of an intervention or programme (Duignan 2009a). These are shown in Table 2.

Table 2 Types of evaluation and their purposes

Type of evaluation	Purpose of evaluation	When in project/ programme cycle
Ex ante	To assesses likelihood of achieving the intended results of a programme/ intervention	Prior to start
Formative	To improve programme or intervention performance and management.	Any phase, but more common near beginning
Develop-mental	To develop an innovation; changing the programme model itself in response to changed conditions and circumstances.	Interaction over project life
Summative	To assess change /effects brought about by a programme, intended or unintended, positive or negative.	Usually at the end of a programme/project phase; sometimes at mid term.
Impact	To measure the extent to which planned and observed changes in outcomes and longer term impacts can be attributed to the intervention. Requires the construction of a counterfactual	Usually ex post.

Developmental evaluation (not to be confused with the evaluation of development) is where an evaluator working as part of a team facilitates a process of discussion around evaluative questions and data-based decision-making in the developmental process. This type of evaluation is seen as relevant to the complex, dynamic and unpredictable scenarios of RAS (Donaldson et al. 2010). In contrast to summative evaluation which measures performance and success against predetermined goals and defined outcomes, developmental evaluation develops measures and tracking mechanisms as outcomes emerge.

These different types of evaluations are also linked to potential uses. One important distinction between evaluation objectives can be drawn between the intended uses of evaluation findings and the intended process use (Hoffmann et al. 2009), which are ultimately linked to the question whether the main aim of the evaluation is to prove or to improve (Rennekamp and Arnold 2009). The table below summarizes the uses of evaluation based on this distinction.

Table 3: The uses of evaluation

...to use findings, results	...to use the process
To see what has been achieved and to justify funding by clients	To improve communication, information and the relationship between clients and extension organizations
To identify strengths and weaknesses, learn from errors, feed back information into planning and improve the extension approach	To create an environment of critical self-reflection and a culture of learning
To ascertain cost effectiveness	To empower clients (encourage clients through 'downward accountability')
To generate knowledge and share lessons and new concepts inside the system	To generate knowledge together and share lessons and new concepts with those outside the system
To influence policies and sectoral priorities	

Source: Adapted from Hoffmann et al. 2009

It is important to note that these two purposes are not mutually exclusive, as those commissioning evaluations (whether donors, governments, NGOs etc.) will continue to request cost effectiveness and impact evaluations whilst recognizing that evaluations for process uses are increasingly important to respond to the growing complexity of extension situations. However, there are likely to be trade- offs and possible tensions between these two aspects of evaluations.

Important factors underlying any evaluation include the policies, requirements and associated guidelines of the commissioning organization. These perspectives on the purpose of evaluation play an important role in determining how the evaluation is designed and conducted. For example, a review of the stated purposes of evaluation of some of the major international development donors[3] reveals several common factors:

- An emphasis on **objectivity and independent assessment of the performance** of programmes and projects, the extent to which activities have been performed as intended and expected results have been achieved together with assessment of their efficiency and effectiveness. Qualities such as reliability and rigour are valued. The assessment is seen as contributing to enhanced relevance, effectiveness and improved performance.
- The importance of evaluation as a means **to demonstrate and reinforce accountability** in the achievement of its objectives by assessing the impact of funded activities. It is seen as generating impartial and unbiased information and documentation to publically account for aid expenditure, processes and outcomes. It can support mutual accountability between stakeholders in development interventions.
- Systematic **identification and dissemination of the lessons learned from experience,** both successes and shortcomings, *"to tell it the way it is"* and show what is working and what is not. Some donors see this as strengthening their institutional learning culture and that of their development partners, which can lead to improved design and delivery of future initiatives and activities. It can generate general knowledge on development which can contribute to maintaining and improving the quality of aid.
- A related purpose is **to improve management and decision making** by providing information for managing projects and programmes and evidence on which to base decisions by development partners as well as lessons to guide future decisions.

[3]EC, World Bank, IFAD, UNDP, FAO and bilateral donors – CIDA, GTZ, NORAD, DANIDA, SIDA, IDRC.

Other objectives less frequently cited by donor agencies were to enhance capacity in projects and programmes and to assess the sustainability of results, indicating that these dimensions have received less emphasis. The relative emphasis on accountability versus learning within evaluation commissioning organizations will influence the formulation of the objective for a specific evaluation and hence the evaluation questions, criteria and approach.

Users, uses and participation

A further important aspect for evaluation planning is the consideration and involvement of users. Users in the context of extension evaluation can range from donors, government, project staff, and the development community to households and individuals. Barker (2007) emphasized that stakeholders are essential to extension initiatives as they provide the link between extension priorities or purposes and the target community. Patton (2008) suggests that stakeholder involvement is also crucial for evaluating extension initiatives and emphasizes the role of utilization-focused evaluation in this context (see box below).

A particular strength of utilization-focused evaluation is a high degree of situational responsiveness, which guides the interactive process between evaluator and primary intended users (Patton 2002b). This is a key characteristic which is valuable in highly complex situations, where any evaluation approach needs to be based on the local conditions and requirements.

Box 1: Key elements of utilization-focused evaluation

Utilization-Focused Evaluation (U-FE) begins with the premise that evaluations should be judged by their utility and actual use; therefore, evaluators should facilitate the evaluation process and design any evaluation with careful consideration of how everything that is done, from beginning to end, will affect use. 'Use' concerns how real people in the real world apply evaluation findings and experience the evaluation process. Therefore, the focus in utilization-focused evaluation is on intended use by intended users. Since no evaluation can be value-free, utilization-focused evaluation answers the question of whose values will frame the evaluation by working with clearly identified, primary intended users who have responsibility to apply evaluation findings and implement recommendations.

Source: Patton (2008)

Findings from the analysis of extension case studies (Pound et al. 2011) suggest that none of the extension projects evaluated has implemented a stakeholder analysis as part of the evaluation process. Furthermore, it was found that often only the objectives were clearly stated, with little attempt to define use and users of the evaluation (Pound et al. 2011), which was also reflected in a lack of specified dissemination strategies for the evaluation findings and recommendations.

The importance of a strong user focus leads to the question of user involvement. In the past the main evaluation mode was that of expert evaluations, which could be either *externally* or *internally* implemented. The table below summarizes the advantages and disadvantages of expert modes of *external* and *internal* evaluations.

Table 4: Advantages and disadvantages of external and internal expert evaluations

External	Internal
Can take a fresh look at the programme	Knows the programme only too well
Not personally involved, so it is easier to be objective	Finds it hardest to be objective
Is not part of the normal power structure	Is a part of the power and authority structure
Gains nothing from the programme but may gain prestige from the evaluation and the organisational or donor connections which may affect response	May be motivated by hopes of personal gains
Trained in evaluation methods. May have experience in other evaluations. Regarded as 'expert' by the programme.	May not be specifically trained in evaluation methods. Has no more training than others in the programme.
An 'outsider' who may not understand the programme or the people involved	Is familiar with and understands the programme, and can interpret personal behaviour and attitudes
May cause anxiety as programme staff and participants are not sure of motives	Known to the programme, so poses no threat of anxiety or disruption. Final recommendations may appear less threatening.

Source: Adapted from Hoffmann et al. (2009)

However, more recently, evaluations conducted by mixed teams of external and internal evaluators have become increasingly common. A further variant is the form of an internal, but independent evaluation run by an independent evaluation department within an organization.

A third category could be seen in p*articipatory* evaluations, which do not solely rely on 'experts' but rather on a range of different stakeholders present in an evaluation context (Ljungman et al. 2005). Table 5 compares expert evaluation approaches with participatory evaluation approaches.

Table 5: Expert evaluation and participatory evaluation approaches

	Expert Evaluation	Participatory Evaluation
What	Information required by funding agencies and other external stakeholders Standards of performance defined by experts, often with reference to formal goals and objectives	To empower participants to initiate, control and take corrective action Community members and other participants set their own standards of success
Who	Expert evaluators in consultation with stakeholders	Community members in collaboration with project staff and external facilitators
How	Evaluators control data gathering and analysis. Scientific criteria of objectivity. Outsider perspective. Long feedback loops.	Self evaluation. Collaborative processes of data collection and analysis. Simple qualitative and quantitative methods. Immediate sharing of results.
When	Mid-term, completion, ex-post	Continuous and iterative. Not sharply distinguished from monitoring.

Adapted from Narayan (1993)

The degree of user involvement and the specific evaluation phases in which different users are involved in the evaluation process are important aspects which influence the evaluation process. Looking back at the complexity of many extension situations, it becomes obvious that there is a range of different users and other stakeholders who could contribute to and benefit from involvement in different phases of the evaluation process. Whereas in simple extension situations it might be quite obvious who should be involved, complex extension situations require a careful process of stakeholder identification. A number of challenges need to be confronted in these complex situations, ranging from the management of conflicting interests and perspectives, to managing power inequalities and different motivations for involvement.

Ideally, planning for impact assessment should begin at the early stages of a project or programme and include those with evaluation experience. This will help to ensure that the basis of comparison for impact assessment is appropriately designed and clearly understood (whether based on random assignment, matched intervention and 'control' groups or comparison over time) and that baseline data are collected. This requires stakeholder participation. Stakeholder participation in the design and planning stages of evaluations can help to ensure the evaluation is both realistic in scope in relation to resources and is of sufficient rigour and will provide the expected information. It can identify additional partners for an evaluation if particular specialist areas are needed.

Evaluation standards and criteria

This section initially examines the existing basic principles of evaluation and the evaluation standards operating among the major international and bilateral donor agencies, since these are influential in evaluation practice with respect to their funded programmes. It then discusses the relevance of these criteria for evaluation in the context of regional-

ly and nationally commissioned evaluations of RAS programmes with different channels of accountability. For example, for complex extension programmes the criteria would need to span a range of actors (knowledge brokers, intermediaries, entrepreneurs, producers, traders etc) as well as the context of demand and the policy and legal context. The challenge is how to capture and assess complex social, behavioural and institutional change.

Evaluation Principles and Standards among major international donors

An important document setting out the DAC Principles for Evaluation of Development Assistance was produced in 1991 (OECD 1991). This was extensively reviewed in 1998 (OECD 1998). The 1991 document defines evaluation as:

"an assessment, as systematic and objective as possible, of an on-going or completed project, programme or policy, its design, implementation and results. The aim is to determine the relevance and fulfilment of objectives, developmental efficiency, effectiveness, impact and sustainability. An evaluation should provide information that is credible and useful, enabling the incorporation of lessons learned into the decision-making process of both recipients and donors."

It defines the purposes of evaluation as: to improve future aid policy, programmes and projects through feedback of lessons learned, and to provide a basis for accountability, including the provision of information to the public. Other purposes concern the promotion of process-oriented and participatory approaches; evaluation as a management tool for decision making and as a way of increasing knowledge about development assistance and its possibilities and limitations as an instrument of economic and social change.

The document elucidates a number of basic principles of evaluation.

- ***Impartiality and independence*** *of the evaluation from the process concerned with policy-making, and the de-*

livery and management of development assistance. This was seen as contributing to the credibility of evaluation and the avoidance of bias, providing legitimacy, and reducing the potential for conflict of interest. This principle implies an organizational structure which clearly separates evaluation and operational responsibilities. The 1998 review acknowledged that if institutional lesson learning is an important objective, the principle of independence has to be balanced with promotion of participation, relevance and ownership of the evaluation products and recommendations.

- ***Credibility and transparency*** are closely linked to impartiality and independence. Transparency is helped by an open process including participation of recipient countries and supported by reporting both successes and failures, making results widely available, distinguishing between findings and recommendations and including information in a way that does not compromise sources.
- ***Usefulness*** – for evaluations to be useful, the findings must be perceived as relevant. Their presentation should be clear, concise, accessible and timely if they are to have an impact on decision-making and should reflect the different interests and needs of the many parties involved in the programme. A key test of usefulness is whether action is undertaken according to recommendations. Other issues concerned the need for practical, action-oriented findings, rigour in data collection and understanding of the institutional, historical, or local context.
- ***Participation of funders and stakeholders*** – whenever possible, both funders and programme stakeholders should be involved in the evaluation process, with the issues of concern to each partner represented in the terms of reference. This provides an opportunity for learning by doing and strengthening skills and capacities in the recipient countries. The 1998 review found that evaluation was still predominantly a donor-driven process mainly due to time constraints and high costs, communication problems, and, occasionally, political obstacles.
- ***Donor cooperation*** – The standards encourage donor collaboration to avoid duplication of effort, to enhance joint learning, to develop evaluation methods and improve mutual understanding of procedures and approaches, share reports and information, and improve access to evaluation findings. Joint evaluations also reduce the administrative burden on the recipient. However, there have been limited examples of successful collaboration and commitment to replace individual agency evaluations, while joint evaluation has been slow to develop. This issue has become more urgent with the development of multi-

donor support programmes (e.g. sector support programmes, co-financing arrangements etc.)

- ***Evaluation programming*** – This principle advises on the development of an overall plan for the evaluation including the activities to be evaluated, priorities and timetable, taking into account the requirements of management and policy makers. These decisions should involve the users of the evaluation outputs. In 1998, users of evaluation reports suggested that more transparency in design and implementation of evaluation would be attained by using logical frameworks, indicators, and "success rating" systems. Guidelines and standards for evaluations, particularly for project evaluations, are common, although their application is not always consistent.
- ***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. It should examine the project/ programme rationale and relevance of objectives and whether they are realizable; whether they have been achieved and the major factors influencing this. Other criteria are the impact and effects of the project/programme, both intended and unintended and the positive or negative impact of external factors. The evaluator should assess whether the impacts justified the costs and whether there are better ways of achieving the results. Sustainability of the achievements in the longer run is a further issue. While recognizing the difficulty of attributing responsibility for outcomes, the principles suggest that an attempt to establish causal relationships must be made. Feedback from member countries recommended a stronger emphasis on results and performance rating and coverage of specific programme interests, e.g. gender, environment and poverty as well as adaptations for evaluations of new forms of assistance, e.g. sector-wide assistance programmes, national development programmes, etc.
- ***Reporting, dissemination and feedback*** – this principle noted that "dissemination and feedback must form a continuous and dynamic part of the evaluation process. It emphasizes clear, jargon-free reporting, the inclusion of an executive summary a profile of the activity evaluated, a description of the methods used, the main findings, lessons learned, conclusions and recommendations. Systematic dissemination and feedback through different channels, formal and informal, would help ensure improved planning and implementation of policy and programmes. This requires resources and senior management support. The 1998 members'

survey indicated that little attention was being paid to the demand, use and usability of evaluations, and differentiation according to different audiences and lesson sharing. A more effective approach to dissemination was needed; for example, through active promotion by the members in their agencies in interaction with officials and professional groups in the partner countries.

A further key document by the OECD Development Assistance Committee's Network on Development Evaluation is the Quality Standards for Development Evaluation (OECD 2010). This document provides guidance on good practice in evaluation, building on the 1991 principles. The standards were produced in draft in 2006 and tested before finalizing in 2010. They aim to improve the quality of development evaluation processes and products, facilitate the comparison of evaluations across countries, support partnerships and collaboration on joint evaluations, and increase development partners' use of each other's evaluation findings.

The standards deal with evaluation processes and products, while reaffirming principles of transparency, independence, credibility, and affirming the importance of evaluation ethics and codes of conduct, respect for human rights and cultural and religious differences, mindful of gender roles, ethnicity, ability, age, sexual orientation, language and other differences when designing and carrying out the evaluation. Furthermore, the standards recommend a partnership approach, including different stakeholders, coordination with national and local evaluation systems and support for capacity development among development partners. They advise quality control through internal and/or external review or advisory panels.

Individual donors' guidelines closely relate to the above principles and standards.

At the European Commission (EC), evaluation is conducted in accordance with a set of evaluation standards (European Commission n.d.) which are expressed as a set of guiding principles that apply to both internal and external evaluations or combinations. Different standards apply for ex-ante evaluations and impact assessments. The evaluation principles of European bilateral donors, for example, GTZ (GTZ 2006, 2007; Reuber and Haas 2009), DANIDA (DANIDA 2006a and 2006b), Sida (Sida 2004) and NORAD (NORAD 2006) closely follow the DAC principles and standards. NORAD (2006) notes in addition the increasing importance of sector programs and budget support and donor harmonization and the need to adapt to these changes with more joint evaluations and closer cooperation within the OECD/DAC evaluation network.

Sida is committed to using country systems for monitoring, reviews and evaluations whenever feasible and appropriate.

Similar principles inform the evaluation practice of other donors.

United Nations agencies follow the United Nations Evaluation Group standards (UNEG 2005) which identify 13 norms of evaluation, including independence, transparency, consultation and follow up, evaluation ethics and contribution to knowledge building.

The evaluation methodology in use at the IFAD Office of Evaluation (IFAD 2009) is based on the principles set out in IFAD's evaluation policy, approved in 2003 (IFAD 2003). The manual builds on internationally recognized evaluation practices and criteria. Similarly, UNDP's evaluation policy seeks to increase transparency, coherence and efficiency in generating and using evaluative knowledge for organizational learning and effective management for results, and to support accountability (UNDP 2009). These norms and standards are largely in line with the standards of the OECD-DAC.

The World Bank distinguishes between self-evaluations by the units responsible for particular programmes and activities and independent evaluation by the Independent Evaluation Group (IEG). Both types of evaluation are subject to the same principles as the DAC principles above – usefulness, credibility, transparency, and independence (World Bank n.d.).

The IDRC Evaluation Unit follows standards of utility, feasibility, accuracy, and propriety (IDRC n.d. and 2007). They emphasize the importance of participation by relevant users for the production of relevant, action-oriented findings. IDRC also emphasizes capacity strengthening, stating that evaluation processes should develop capacity in evaluative thinking and evaluation use among IDRC managers, programme staff, or project partners – building an evaluation 'culture'. Specific strategies can be built into evaluations to foster this.

The principles which guide evaluation among charities and foundations are less readily accessible. The W. K. Kellogg Foundation (2004a) has detailed information on evaluation and sees the guiding principles as follows: strengthening projects, using multiple approaches, addressing real issues, and a participatory and flexible process.

In summary, the development of evaluation principles and standards by the major funding agencies contributes to a broad framework for evaluation practice, applicable to programmes in different sectors. However, the particular emphasis on the different

standards will depend on who is driving the evaluation and with what purpose.

Evaluation criteria

Understanding of the evaluation principles and standards – whether formalized as discussed above, or implicit in organizational norms and practices – forms the basis on which evaluation objectives are defined and specific evaluation questions developed for particular programmes or projects. Evaluation practice varies according to whether standardized evaluation criteria are used and adapted or whether the evaluation criteria and associated questions are developed as part of the evaluation process, for example linked to the analysis of the intervention logic.

The most widely used criteria are linked to the DAC principles and standards (EuropeAid 2006). The DAC evaluation criteria – relevance, efficiency, effectiveness, sustainability and impact[4] – represent different dimensions or perspectives for assessing the performance of a project.

- **Relevance** – the extent to which the objectives of the development intervention are consistent with beneficiaries' needs and problems, country needs, global priorities and partners' and donors' policies; whether the objectives continue to be relevant.
- **Effectiveness** – the extent to which the objectives of the development intervention were achieved, or are expected to be achieved, taking into account their relative importance.
- **Efficiency** – examines how resources – inputs, funds, expertise, time – have been converted to results and whether the results were achieved at a reasonable cost.
- **Sustainability** – the extent to which the benefits from a development intervention continue after major development assistance has been completed and the probability of continued long-term benefits.
- **Impact** – the positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended.

[4]To these, the EC add coherence/ complementarity and community/value added.

The additional EC criterion of coherence/complementarity is important for evaluations of complex programmes – examining the 'fit' of the programme with donor and country policies and priorities. It can also refer to the logic of the relationship between activities, outputs and results, examining whether there are internal contradictions which affect the results.

The OECD DAC criteria are utilized, with minor modifications, by many donors and development organizations, including the World Bank (World Bank n.d.), IFAD (2009), UNDP (UNDP 2009), FAO (2007 and 2009), GTZ (2006 and 2007), NORAD (2006), and DANIDA (2006b). The main area in which additions have been made concerns institutional development and partnerships. The World Bank adds 'institutional development impact' and 'bank and borrower performance'. IFAD includes 'innovations and performance of partners', and FAO, 'institutional strengths and weaknesses, including institutional culture and inclusiveness of process'. CIDA includes partnership in their development results and success factors (CIDA 2004 and 2006).

IDRC's approach to evaluation is utilization-focused and does not advocate any particular evaluation content, model, method, or theory. Rather, the primary intended users of evaluations should select the most appropriate focus, methodology and approach. A key outcome of the evaluation process is informed social participation, making a contribution to better governance, and transparency (IDRC 2007).

Interestingly, the meta review of evaluation case studies conducted as part of the same commission as this literature review (Pound et al. 2011) found that about half of the selected cases referred to some or all of the five OECD DAC criteria, but in most cases these were not linked to the evaluation questions. Relevance and efficiency were less frequently mentioned than the other criteria.

These are broad criteria for evaluation. There are several important methodological steps in moving towards developing the detailed questions for an evaluation.

Impact evaluation

An important debate relevant to evaluation of RAS concerns the meaning and expectations of 'impact evaluation'. The term is often used broadly to refer to an analysis of outcomes and long-term effects of an intervention (White 2009a). For example, the DAC definition of impact is 'positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended'. Any evaluation which refers to impact (or outcome) indicators is thus, by definition, an impact evaluation.

However, increasingly, 'impact evaluation' is being defined more specifically to refer to approaches which address the question of how outcomes such as participants' well-being would have changed if the intervention had not been undertaken (White 2009a). This definition is about *attribution* of changes in the outcome to the programme or intervention. Reasons for interest in such approaches are partly driven by policy concerns and the need to demonstrate cost effectiveness and justify investment.

To analyse attribution, there must be a means of comparing the actual changes brought about by the programme with the situation if the programme had not taken place (the *counterfactual*). This can be done in different ways; for example, comparing the situation before and after an intervention, or comparing groups who were targeted by or participated in the intervention with similar groups who were not.

Because of the interest in establishing causality and the need to exclude influences which could affect results, but which are independent of the programme, some evaluators assert that analysis of attribution requires an *experimental* or *quasi-experimental* design involving the construction of a rigorous counterfactual and, where possible, the use of quantitative measures of impact.

Counter arguments cite the complexity of rural social structures and interaction (particularly in the case of RAS) considering that there are far too many variables and possibilities emerging and interacting dynamically to conceptualize simple counterfactuals (Patton 2010). A further risk in the context of RAS evaluation is that attention to the broader goals of RAS – for example, institutional change, empowerment, gender equity, strengthening social capital and enhancing participation – would be diverted in favour of those which are more easily measurable; for example, diffusion and adoption of technologies.

The implication of the above is that the appropriate approach and methods in evaluation depend greatly on what kinds of questions are being asked and what type of analysis is required. There is a growing suite of methods, both qualitative and quantitative, and the challenge is to select the best methods and combinations of methods for the purpose and resources available.

Analysing the intervention logic or programme theory

There is a broad consensus that the reasoning behind a programme, project or intervention needs to be clearly understood. The evaluation literature and available evaluation guidelines make a number of suggestions which link the development of evaluation criteria to a theory of change or intervention logic. This relates to how particular interventions or activities bring about certain results and what those results will lead to. Other terminologies for a similar conceptualisation are 'impact pathway', 'outcomes model' and 'chain of results'. A theory-based approach can provide a framework for an evaluation (White 2009b). This can be conceptualised as a series of cause and effect relationships linking inputs and activities to expected results[5] and impacts and to overall goals. This is often captured in a project or programme logical framework with associated indicators, time-bound targets and assumptions, which are then used as the basis for evaluation. Some of the shortcomings of logical frameworks and their application are well known, but these are particularly pertinent when it comes to complex programmes or when they are used as the single tool.

- It is difficult to encapsulate all the elements of a complex programme in a logframe, unless it has different interrelated levels. The logical hierarchy makes it more difficult to represent multiple parallel processes – such as programmes at different locations, interactions between different outputs and outcomes, and the different roles of multiple stakeholders. Complex statements can make the logical framework unreadable.
- The assumptions column is often considered 'residual' and important external factors that condition or limit project implementation and results are given inadequate attention, although they may be issues that the project should seek to influence.
- Responsibility and agency – who will do what among multiple partners – is not always clear.
- Logical frameworks are not always 'owned' by the project team. While some may be a product of participatory planning programme processes (e.g. ZOPP - objective oriented project planning) they are often prepared as a separate activity by a consultant to meet donor requirements, and are not used by the project team as a management and monitoring tool (Bakewell and Garbutt 2005).

[5]The OECD DAC criteria utilise the concepts outputs, outcomes and impacts

- The logframe identifies planned results but tends to be inflexible with respect to incorporating changes that were not originally planned or unintended effects.

Use of the logical framework in evaluation is often emphasised where the main driver of the evaluation is upwards accountability to the donor (Bakewell and Garbutt 2005). Its advantage is to focus the enquiry on whether the intervention is delivering the outputs and impacts as proposed (or 'contracted'), and to examine how outcomes and goals have been delivered that can challenge the programme logic and question whether the project is in fact 'doing the right things.' One consequence is that this focus on outcomes and impact may come at the expense of understanding and learning from the processes of project implementation – whether things 'were done right' (Bakewell and Garbutt 2005). However, in practice, the upper levels of logical frameworks are often neglected and their indicators not taken very seriously, especially where the logframe is seen as a donor-imposed requirement. Hence the focus tends to be more on project implementation and whether 'things were done right'.

The use of logframes as the single basis for evaluation can also be problematic when the actual objectives and dynamics are determined by the playing out of stakeholder interests (Gasper 2000).

Where a logical framework does not exist, the intervention logic can be reconstructed using statements of objectives and anticipated results and impacts. This may include multiple stakeholders and be done as a participatory activity, or may simply be conducted by the evaluation team based on project documentation and verified with project stakeholders. The process should clearly indicate what was stated in documentation and what additional results have been added. It can be presented as a diagram or tables. The evaluation questions are then developed from the intervention logic.

'Theory of change' approaches or theory-based evaluation (Chen 1990; Weiss 1995; White 2009b) are similar to logic models in defining how and why the programme works and what transforms actions into interim and long term outcomes. They address processes in more detail, generally involving an examination of the context, assumptions, and preconditions and presenting them in a more interlinked visual format. These approaches evolved in the charitable sector working on community development programmes and other complex initiatives (W. K. Kellogg Foundation 2004b). While logic models have been used in extension evaluation, there is potential for the more flexible use of programme theory to help in specifying causal connections, focusing attention on certain possibly weaker parts of the

model, and identifying critical mediators that are necessary to make change happen (Braverman and Engle 2009; Funnel and Rogers 2011).

A different approach, from an NGO, is ActionAid's Accountability, Learning and Planning System (ALPS). This is a set of minimum requirements for planning and accountability which describes the reviews and reports required within the organisation and a set of principles describing how these requirements should be met. These are closely linked to ActionAid's mission and strategy, intended to promote greater accountability to the people on whose behalf money is raised, gender equity, transparency, and empowerment of the poor (Guijt 2004, 5).

More strongly actor-oriented frameworks are evolving. An example is Outcome Mapping (Earl et al. 2001) which focuses on people and organizations rather than assessing the products of a programme. It starts by establishing a vision of the human, social, and environmental betterment to which the programme hopes to contribute and then focuses monitoring and evaluation on changes in behaviour, relationships, actions, and/or activities of the people and organizations with whom a development programme works directly and has opportunities for influence. These latter are termed *'boundary partners'*.

This formulation is useful for monitoring and evaluating changes in relationships, and how such changes can be brought about. The methodology is helpful in addressing the challenge of monitoring and evaluation within a complex multi-stakeholder context where changes in relationships, processes, attitudes, capacities and quality of communication are important objectives.

Network models (www.mande.co.uk/networkmodels.htm) are useful where there are many actors (people and /or organisations) who are fairly autonomous and where there is no single authority directing them, and where programmes have multiple or competing objectives, or where there complex connections between actors and outputs. This is particularly the case in pluralistic RAS where different organizations are active in the same rural space.

An example is a study from Kenya (AED/USAID n.d.) which used social network analysis as a participatory planning tool for dairy feed system development. Information on relationships and information exchange was mapped and discussed, identifying strengths weaknesses and gaps in the network relationships. Collaborative actions were monitored and changes in the network structure mapped. This creates a foundation for collaborative evaluation of innovations in a single area.

The Innovation Systems Framework also offers a useful framework for analysis, monitoring and evaluation of networks of organizations in extension (Hall et al. 2003; World Bank 2006) and links localized networks to a broader enabling national and international policy and trade environment. Other methods of developing criteria for evaluation where these have not been specified as part of a project design and M&E system (for example, in the evaluation of some humanitarian interventions in conflict situations) include the 'Most significant change'(MSC) technique (Davies and Dart 2005). This is a form of participatory monitoring and evaluation in which many project stakeholders are involved both in deciding the sorts of change to be recorded and in analysing the data. It contributes to evaluation because it provides data on impact and outcomes that can be used to help assess the performance of the programme as a whole. The data are 'significant change' stories collected from the field level, and the systematic selection of the most significant of these stories by panels of designated stakeholders or staff. These are shared and are the basis for discussions about the value of these reported changes and programme impact. MSC does not make use of pre-defined indicators, especially quantitative ones.

Appreciative inquiry is a technique for creating an organisational vision, aligning groups, and building cultures (http://appreciativeinquiry.case.edu/intro/definition.cfm). It searches for the positive in people, their organizations, and the relevant world around them. It does this through a "4D" process to 1) Discover the "best of what is"; 2) Dream "what might be"; 3) Design "what should be", and 4) create a Destiny based on "what will be". AI can be used in evaluations to develop a programme logic model, clarify the evaluation purpose, identify stakeholders, determine key questions, and develop measures and indicators and an evaluation plan (Preskill and Coghlan 2003).

Evaluation questions

Evaluation questions provide a transparent focus for the evaluation and should reflect the priority issues. The choice of questions is particularly critical for evaluations of complex multidimensional programmes – a narrow focus allows more targeted data collection and in-depth analysis, but may risk missing important factors, especially if the intervention logic has been challenged. Hence the recommendation in the DAC evaluation standards document that relevant stakeholders, including the users of the evaluation report, have the opportunity to contribute to evaluation design, including identifying issues to be addressed and evaluation questions to be answered. Question areas for inclu-

sion might concern particularly innovative aspects of a programme which interest stakeholders or topics that will inform a decision or a policy debate.

The evaluation questions relate to the criteria for evaluation and the programme or project logic. The EC evaluation guidelines suggest questions, and associated subquestions should be linked to the DAC criteria (see example in section 10), exploring the extent to which the project activities have led to the intended results (effectiveness), whether the results have been achieved at reasonable cost (efficiency), and whether these have actually been relevant to local needs and with what levels of impact and prospects of sustainability (EuropeAid 2006; European Commission 2004). The guidelines suggest each question should focus on one evaluation criterion as they will require different approaches to collecting evidence. Translating the generic criteria into specific questions is a challenge in itself, but other questions, not directly linked to the intervention logic are also needed – concerning unexpected impacts and their benefits or lack of benefits and cross-cutting issues such as environment, gender, good governance and human rights.

As part of assessing the contribution and the rationale of the project or programme, the questions may also explore what changes would have taken place had the project/programme not been in place or had it worked with different partners. This means considering whether the evidence could be supported by alternative explanations.

For projects with logical frameworks containing indicators and targets, the assessment criteria for the specified results should be explicit. Where these do not exist, assessment criteria, targets and indicators have to be developed and accepted for consistency, transparency and feasibility of data collection.

Evaluation approaches

This section discusses approaches to evaluation in the sense of the overall orientation toward the evaluation, including the underlying value orientation, purposes and methods. Approaches to evaluation are related to the purpose, uses and users of the evaluation as discussed in section 4 above, which are in turn related to the principles, standards and expectations of those commissioning evaluations and those involved in conducting them.

The different purposes of evaluation were presented in Table 2 above. Different purposes of evaluation will be relevant for a programme according to its strategic needs. For example, while formative evaluation to ensure that the programme

is well-developed is clearly important in the early stages of programme development, it can be useful at points along the programme implementation to improve management performance and programme targeting. Formative evaluation may include the development or revision of the programme logic model, examining stakeholders' views of the programme; researching or reviewing information about needs; sharing experience and consolidating learning within and across the programme; and developing data collection systems for M&E (Duignan 2009d).

In addition to types of evaluation defined by purpose, there are several types of evaluation characterised by a specific focus or approach. *Utilization-focused evaluation,* Developmental evaluation and *theory-based evaluation* have been discussed above. *'Process evaluation'* is an evaluation approach which focuses on understanding programme or implementing organisations' dynamics and decisions – why things happen/ed the way they do/did. It can complement other types of evaluations by helping to interpret the context and the interpersonal, institutional and political processes that influence achievements. It is based on exploration and analysis of stakeholder and participant perceptions about the programme, including documentation and communication. It can be combined with other approaches to provide insight into *why* things happened the way they did.

Other types of evaluation include:

- **Empowerment evaluation** – emphasizes that the evaluation process and methods should be empowering to those who are being evaluated (Fetterman et al. 1996, 2004).
- **Stakeholder evaluation** – looks at the different perspectives of all of a programme's stakeholders (where stakeholders are those who have an interest in the programme) (Greene 1988).
- **Goal-free evaluation** – in which the evaluator's task is to examine all of the outcomes of a programme, not just its formal outcomes as specified by programme planners in its objectives (Scriven 1972).
- **Constructivist/naturalistic or fourth-generation evaluation** – emphasizes the qualitative uniqueness of programmes and is a reaction against the limitations of quantitative evaluation approaches (Guba and Lincoln 1989).
- **Experimentalist evaluation** – the traditional quantitative approach to evaluation experiments or quasi-experiments which puts a particular emphasis on their use as ways of providing evaluative information that can be used to attribute changes in outcomes to particular interventions (Cook and Campbell 1979). Similar to some interpretations of impact evaluation.

- **Strategic evaluation** – an approach which emphasizes that evaluations should be driven by the strategic value of answering key sector information needs rather than by just focusing on individual programmes (Duignan 2008).
- **Most significant change** – collects 'significant change' stories from the field. The most significant of these are selected systematically by panels of designated stakeholders or staff to constitute data on impact and outcomes that can be used to help assess the performance of the programme as a whole. (Davies and Dart 2005).

(source: adapted from Duignan 2009d)

Summative evaluations are about assessing the changes brought about by a programme, intended or unintended, positive or negative. The difference between this approach and the specific requirements of impact evaluation, which is concerned with measuring and attributing impact, have been discussed above. While both types of evaluation are concerned with identifying changes that can be plausibly attributed to the programme, impact assessment does this by reference to a 'counterfactual' ('what would have happened to beneficiaries in the absence of the programme'), (see http://en.wikipedia.org/wiki/Impact_evaluation).

Box 2. Examples of designs (Duignan 2009d; EuropeAid 2006)

- Participant and key informant judgement design – using participatory methods, interviews, focus groups, case studies, and an analysis of beneficiaries affected by the project. This relies on constructing 'before and after' comparisons or identification of significant change (for summative evaluation).
- Quasi-experimental design – comparing groups receiving the intervention with similar groups who did not. However, systematic pre-existing differences between the two groups might bias the findings. Matched comparison design involves the use of a control group that matches as closely as possible the characteristics of the 'intervention' group – either through propensity score matching or using a multivariate regression approach. This method often involves the use of large-scale sample surveys and sophisticated statistical analysis.
- Randomized design (experimental design) – this involves the random assignment of individuals or households either as project beneficiaries or as a control group that does not receive the service or good being provided by the project. This is designed to show causal relationships between certain outcomes and the "treatments" or services aimed at producing these outcomes.

Under the heading of summative evaluation approaches are included a range of methods that rely on qualitative and sometimes quantitative approaches. They often explore participants' and stakeholders' experience and assessments and examine institutional change, incorporating 'before' and 'after' comparisons, or monitor and then evaluate changes over time. These approaches can successfully identify and estimate the extent of change and assess performance, but are not necessarily able to attribute the change unambiguously to the programme intervention.

In contrast, for impact evaluations which are interested in analysing attribution, there has to be a means of comparing the actual changes brought about by the programme with the situation if the programme had not taken place (the *counterfactual*). This can be done in different ways; for example, comparing groups who were targeted by, or participated in the intervention with similar groups who were not. There should be an assessment of whether project participants and non-participants are comparable (usually involving some form of statistical matching) to exclude the possibility that pre-existing differences bias the targeting of programme delivery. This can involve complex experimental designs using control groups and sophisticated statistical techniques. Both types of evaluations are more straightforward where relevant baseline data have been collected, but a good baseline study design is the first step in the design of an impact assessment.

These alternatives raise very different issues, including ethical and resource implications (Duignan 2009b). Decisions on

the appropriate approach for evaluation of complex programmes are particularly challenging. Among the considerations are:

- The acceptability of approaches to different client groups – donors and country partners; what is seen as constituting 'evidence' and the level of rigour required (see below)
- The ethical implications of 'excluding' potential beneficiaries from a development programme in order to construct a 'control' group.
- The main function and purpose of the evaluation – including the relative weight given to accountability or 'proving' impact, or to learning lessons for the future and 'improving' performance.
- The type of programme being evaluated and the degree of complexity. For some evaluation approaches, the criteria necessary to conduct these evaluations limit their usefulness to primarily single intervention programmes in fairly controlled environments (W. K. Kellogg Foundation 2004a).
- The scale of the evaluation, the feasibility and the resources and time required for different approaches.

An example of a decision tool to help in decision-making on impact/outcome evaluation design is given by Duignan 2009c. This looks at the relationship between different evaluation types and the characteristics of an intervention – including the level of control over access to an intervention, ease of discerning causal mechanisms, resource levels, the degree of stakeholders' scepticism, and likelihood of completion. Duignan notes that particular stakeholders may reject the last two designs or some of the others as being inadequate for robustly establishing causality in specific cases or in all cases.

Table 6: Impact/outcome evaluation design selection

Context / Design	Cannot control who gets inter-vention?	Cannot stop control/ comparison group getting any inter-vention?	Intervention applied to all?
True Experiment – randomised pre- and post-intervention evaluation	Not appropriate	Not appropriate	Not appropriate
Time series analysis	May be appropriate	Does not rely on a control group	May be appropriate
Constructed matched comparison group (quasi experimental)	May be appropriate	Not appropriate	Not appropriate (except different time period)
Ex post participatory assessment. Triangulation with other sources.	May be appropriate	No formal control group	May be appropriate
Expert judgement	May be appropriate	No formal control group	May be appropriate
Key informant	May be appropriate	No formal control group	May be appropriate

Adapted from Duignan 2009c and World Bank 2004

It is thus very difficult to make an overall recommendation on the best approach for agricultural service programmes. However, since important dimensions of agricultural service delivery include assessment of changes in social networks, communication and information flows, partnerships, partners capacity etc., it is important that evaluation approaches include methods to explore these dimensions which are not

	Causal mechanisms difficult for key informants/ experts to discern?	Limited resources for impact/ outcome evaluation?	Powerful stakeholders sceptical about interven-tion?	Concern about risk of impact/ outcome evaluation not being com-pleted?
	May be appropriate	May be more expensive	May be more appropriate	Higher
	May be appropriate	May be less expensive	May be more appropriate	Lower
	May be appropriate	May be less expensive	May be more appropriate	Lower
	Not appropriate	May be less expensive	May be less appropriate	Lower
	Not appropriate	May be less expensive	May be less appropriate	Lower
	Not appropriate	May be less expensive	May be less appropriate	Lower

easily amenable to randomised design or control groups and where it is more difficult to be confident about attribution. Hence there is a need to consider combinations of methods in evaluation of extension initiatives and not to retreat from the evaluation of complex programme areas. For example, to understand changes in interrelationships and partnerships, outcome mapping might be used.

Rigour

Braverman and Arnold (2008, 82) define rigor as "a characteristic of evaluation studies that refers to the strength of the design's underlying logic and the confidence with which conclusions can be drawn". Funding agencies, grant reviewers, legislators, and our academic departments desire methodologically rigorous evaluations of extension programmes, that is, evaluations that are technically sound and provide an opportunity to show solid, convincing evidence of a programme's impact (Duniform et al. 2004).

Rigour contributes to evaluation quality, and it can be described in terms of specific elements related to the evaluation's planning and implementation. Those critical elements include the following (see Braverman and Arnold 2008 for a more detailed list):

- Evaluation design: For programme impact evaluations, how well does the design allow us to determine if the programme itself was the cause of positive change in the outcomes?
- Measurement strategies: Will the programme outcomes be measured in a valid, reliable way that provides strong evidence for drawing conclusions?
- Programme monitoring: During the evaluation, are we observing the programme closely enough so that we can describe how it is being delivered, including potential differences between programme delivery sites?
- Programme participation and attrition: Are efforts made to reach those who did not participate or whom the programme did not reach: participants who didn't attend regularly, who left the programme midway, or who received different levels of exposure to the programme? Or does it just measure whoever happens to attend on the day of data collection?

An inevitable dilemma in impact assessment is the need to be rigorous and the need to be comprehensive. Here, 'rigorous' is meant in the narrow sense – in terms of obtaining representative and convincing statistical data – implying a restricted scope. Comprehensiveness is meant in terms of a broad range of technical, educational, social and political impacts. Studies have been designed to be either rigorous or comprehensive but never both (van de Berg 2004).

In complex situations such as community initiatives, or those equally applicable to complex RAS, the conventional hypothetico-deductive, natural science model, which results in high levels of statistical rigour, is ill-equipped to capture the multi-dimensional impacts (W. K. Kellogg Foundation 2004b), leading to an incomplete understanding of the intervention dynamics and system changes.

The consequences are:

- Exclusion or narrowing of the choice of alternative paradigms and associated questions which, if chosen, are seen as being of lesser value.
- Since it is difficult or impossible to apply the dominant evaluation paradigm, such initiatives may not be evaluated at all and evidence of their effectiveness will be lacking. Alternatively, using the standard impact evaluation methods, the range of impacts would not be captured.
- This might prompt the revision of programme activities and design away from the original objectives.

Quantitative measures have a built-in bias toward high-external-input methods which maximise production through potentially unsustainable techniques. If evaluation parameters based on production goals have been pre-set, the empowerment of farmers to change the extension agenda to meet their own needs will be judged as a failure of the project rather than evidence of readiness to absorb feedback. One lesson learned from project experience in the Mekong Delta was how traditional, quantitative targets can rule out the investment required to develop a relationship with the poor (Christoplos 1996). Emphasis on rigorous quantitative and cost-benefit evaluation of the impacts of extension investments may cause the extension service to adopt an elite bias, working with wealthier farmers who are more likely to show greater production increases for lower administrative and logistical costs (Christoplos 1996). In contrast, poorer farmers operating diversified systems, where the production is difficult to aggregate and quantify, are more likely to be ignored.

It is therefore necessary to decide on the 'right' level of rigour for a given situation, based on purpose, actors, etc.

Table 7

Options in evaluation planning: moderate-rigour and high-rigour choices to evaluate a multiple-session, multiple-site parenting skills education programme			
Rigour element	**Moderate-rigour option**	**Higher-rigour option**	**What the higher-rigour option adds**
Evaluation study design	Single group pre- and post-test design	Comparison group design	More confidence that our programme was the cause of positive change (if indeed positive change occurs)
Measurement – knowledge gain	Participants' self-ratings of how much they learned about good parenting	Valid, reliable test of what people actually know about the programme's content	Being able to make more authoritative statement about what people really know (and don't know) after participating in the programme
Measurement - behavioural change	Participants' intentions (at end of class) to change their parenting behaviours	Six months after programme, self-report surveys of participants' current parenting behaviour	More confidence in stating that the programme has resulted in actual behavioural change
Programme delivery monitoring	Observe one session per delivery site, or interview programme leader to determine what content was covered	Observe multiple sessions at each delivery site to get a detailed picture of programme delivery	Ability to explain, rather than speculate, about why delivery sites may differ from each other in effectiveness
Programme participation and attrition	Give survey to only those participants who attend the final class session	Programme team determines beforehand what minimum number of sessions should count for programme participation, and makes attempt to survey an appropriate sample of participants who meet that attendance standard	More comprehensive understanding of the programme's full audience, rather than a convenience sample of people who attended on a given date

Braverman and Engle (2009)

Other points to consider in evaluation planning are how the programme or intervention was targeted to individuals, households, communities or other stakeholder groups – whether randomised assignment, or individual participant self selection, or selection on the basis of programme or community determined criteria.

There are principles and approaches that can ensure greater rigour in the use of non-experimental and participatory methods. Triangulation of data collection is an important principle in using qualitative methods. It is important to ensure that a thorough stakeholder analysis is carried out and the interactions based on a selection of these groups (Guion 2002). Triangulation of methods can add further rigour by collecting similar data using different methods, both quantitative and qualitative, and by comparing conclusions. Environmental triangulation involves the use of different locations, settings and other key factors related to the environment in which the intervention took place and which may influence the information received during the evaluation (Guion 2002).

Attribution

Attribution analysis aims to assess the extent to which the observed change can really be attributed to the evaluated intervention. Some writers consider this to be synonymous with impact evaluation. One challenging aspect of evaluating new models in agricultural service provision is that they often represent different combinations of change in governance structures, capacity, management and advisory methods. Moreover, which combination is most suitable for a given situation depends on a number of frame conditions, such as the type of farming system, socioeconomic conditions, and state capacity (Birner et al. 2006). Hence it is inherently difficult to establish which factors account for observed changes in outcome if a new model is introduced (Anderson 2007).

In terms of the outcomes and impacts of agricultural service provision, there is a high level of risk and unpredictability associated with factors outside the control of advisory service providers. Many of these factors, for example climate variability, political instability, and market price volatil-

ity, can have serious negative direct and indirect effects on the operation of otherwise well-designed programmes.

The level of attribution is related to the choice of evaluation design discussed above and the use of evidence to construct an 'argument' about the effects of the programme interventions. Comparing evidence to programme theory diagrams (from the original programme design or constructed by the evaluation team) can assist in this; however, alternative explanations should also be considered

For complex RAS, isolating the impact or contribution of one service or one element is always challenging and sometimes impossible. It certainly helps when the issue is considered at project design stage.

Box 3: African Farm Radio Research Initiative (AFRRI) – use of a counterfactual in a project design

The Farm Radio Research Initiative aims to assess the effectiveness of farm radio in meeting the food security objectives of rural farming households in Africa. Working with five partner radio stations in each of five African countries – Tanzania, Uganda, Mali, Ghana and Malawi – AFRRI is producing, broadcasting and evaluating the outcomes of participatory radio campaigns. For each radio station, three different categories of communities were identified as part of the assessment study: the active listening communities, who were those directly participating in project planning, production and evaluation; the passive listening communities that had access to the radio programme, but no involvement with the radio station or the project; and the control communities, those who were not able to listen to the programme at all. A challenge related to maintaining a proper "control" community lay in the fact that the project had no control over information exchange between the active listening communities and control communities, and no influence on changes in radio coverage areas. Comparisons between the active and passive listening communities helped to isolate the effects of hearing the radio programmes on improved rates of adoption, from the effects of direct participation in the project. It also indicates the value-added of more direct face-to-face participatory engagement with listening communities.

Source: African Farm Radio Research Initiative 2009

Required principles for RAS evaluation approaches in high complexity situations

So far we have reviewed the different criteria applied in evaluation and have established the relevance of the DAC evaluation criteria in development aid, which includes support to rural advisory services. Guiding principles for the RAS evaluation process should be the use of the DAC criteria as a framework to structure the evaluation content as well as the key characteristics of a pluralistic and demand-driven extension service.

The key characteristics of RAS based on GFRAS (2010a), Anderson (2007), Hoffmann et al. (2009), and Patton (2008) can be summarised as follows:

- Inclusive: RAS should embrace diversity of users and needs to be suitable for different genders, ethnic groups and different socio-economic categories, with a particular focus on pro-poor inclusiveness. Participation of users is a key element of such service.
- Demand-driven: A diverse range of demands can be expected from diverse user groups. Demand will differ based on situational context (value chain, markets and market trends, pro-poor focus, etc.)
- Pluralistic: The service will need to take into account the co-existence of different delivery agents and sources. Therefore, networking and cooperation are key elements.
- Learning and capacity focused: A key element of the service is to enhance users' knowledge and capacity to access and utilise RAS and to develop additional capacities.
- Adaptive/ change oriented: No single development pathway is likely to be suitable in complex social and environmental conditions. Therefore, the service needs to be adaptive and flexible in order to react to emerging opportunities and needs. Flexibility and practical responses to risks and changes in circumstances are required.

The table below is an example of how the five DAC criteria could be linked to the attributes of emerging RAS services.

Table 8: Agricultural extension evaluation matrix: attributes of RAS by evaluation criteria

DAC evaluation criteria	Inclusive - Participation gender, ethnic, poverty focus	Demand-driven Market-oriented
Efficiency	Ratio of resources used/ poor groups as participants	Proportion of resources used in identifying and addressing demands. Return on resources invested in market development
Effectiveness	How representative are client groups of population?	How well do demands translate into service provision and market access?
Relevance	Are the needs of poorest/ marginalized recognized?	How well are demands understood and assimilated? How have opportunities along the value chain been identified?
Sustainability – technical	Do tools for inclusiveness maintain breadth of demand side?	Are the methods for assessing demand changes robust over time?
Sustainability – institutional	Are the incentives to maintain focus on poorest/ marginalized sufficient over time?	Is demand responsiveness maintained? Are market linkages sustained without programme support?
Impact	What are the livelihood changes attributable to the programme among different social groups?	Have demands been met and benefits gained through increased market access?

This matrix could serve as a framework for the evaluation design; however it does not on its own lead to a specific evaluation approach. The purpose, focus and questions of the evaluation, the resources at hand (time, money, human resources) and the

	Pluralistic delivery agents Partnerships / networking	Learning / capacity strengthening	Adaptive change orientation
	Number and diversity of delivery agents. Frequency and number of farmer contacts	Resources used to enable reflection and analysis of performance	Response time of reactions to changing circumstances – inertia
	To what extent are the delivery agents different in performance?	How well are lessons assessed and absorbed into practice?	Do responses change performance and for how long?
	Is the diversity of delivery agents appropriate to the breadth of demand side?	Are the important lessons learned	Do responses to changed circumstances improve performance?
	Is the turnover in delivery agents maintaining diversity of supply?	Does learning continue?	Do new circumstances challenge/ defy learning?
	How well is the breadth of delivery agents maintained? Are local actors committed to maintain and invest their own resources in these areas over time?	How well is the institutional memory and intelligence managed?	Responsiveness to changes over time
	Change in management and agricultural productivity	How have lessons shared influenced policy-level decisions and investments?	Is there increased resilience to change?

theoretical perspectives are all elements which inform the choice of evaluation approach (Hoffmann et al. 2009).

Strengths, weaknesses and gaps in extension evaluation approaches suitable for complex situations

The Neuchâtel Group has identified the need for common approaches to the monitoring and evaluation of extension activities. Their recommendation is that to improve the analysis of extension schemes, clear frameworks should be offered to evaluators of projects with an extension component. Donor agencies could also devise common reporting procedures (Neuchâtel Group 1999). Although in theory this seems a desirable strategy, it might prove difficult to identify 'best practice' approaches to implement within such frameworks. As we have shown above, there are a range of possible approaches which could be adopted for the evaluation of complex extension activities. In practice, it is more a matter of defining 'best-fit' approaches according to the specific circumstances. However, a set of commonly adopted principles, including the improved and coherent reporting requirements mentioned above, could be an effective way to improve future extension evaluation strategies and facilitate joint learning. In this section we will identify strengths, weaknesses and gaps in current evaluation practice.

Participation and user focus, demand-driven

Growing evidence suggests that stakeholder involvement is a key requirement for successful evaluation practices. Patton (2008) stresses that the evaluation process must discover the key stakeholders, namely those who really want to know something, as this not only increases the chance of usefulness but also of ownership and acceptance. Ownership can lead to improved uptake, whereas the lack of ownership of evaluations delivered externally can explain (at least partly) their lack of influence (Jones and Mendizabal 2010). External evaluations are intended to provide an independent perspective (van de Berg 2004), but it should be recognised that they can also contradict the aim of transferring power to users. Furthermore, they can be costly, and the lack of close association of the evaluators with a programme can hamper sample selection, the choice of parameters, and the interpretation of results (van de Berg 2004). In contrast, internal evaluations are potentially more relevant and comprehensive, especially when primary stakeholders are involved in designing the evaluation (van de Berg 2004). However, other voices point out that direct or indirect stakes in a programme can cause results (e.g. benefits, disadvantages, costs) to be overstated or understated (van de Berg 2004). A possible way forward is mixed-team evaluations, which can combine the benefits of both.

An important aspect to clarify here is the quality of stakeholder participation. Participation in evaluation spans a gradient from complete user (client)-controlled evaluations to initiatives by researchers and development agents or governments to consult users about the results of interventions, to the participation of field workers and researchers in evaluation – as opposed to external evaluations by funding agencies with little focus on user involvement. Lawrenz and Huffman (2003) suggest the use of the following four criteria to determine the extent of user participation: (1) type of evaluation information collected, such as defining questions and instruments; (2) participation in the evaluation process; (3) decisions concerning the data to provide; and (4) use of evaluation information.

Lessons from participatory evaluations in community development projects outlined by Mancini and Jiggins (2008) established that participatory approaches to developing indicators provide opportunities for community empowerment not otherwise provided by top–down approaches (Fraser et al. 2005): they ensure that the indicators are relevant to local situations; they measure what is important to the community; and they are adapted over time as community circumstances change (Pretty 1995; Carruthers and Tinning 2003; Freebairn and King 2003). The involvement of communities helps to build community capacity to address future problems and leads to local action to deal with current problems. These participatory processes, however, have methodological implications; for example, Carruthers and Tinning (2003) and Reed et al. (2005) suggest that the methods used to collect, analyse, and interpret data must be easy enough for active participation by local communities.

On the other hand, community indicators can be very specific to communities – based on their experiences and on the local context. This makes it difficult for the wider extrapolation and comparison of indicators across communities and across landscapes. Because of some of these shortcomings, expert-led or external-led approaches argue the need for generic indicators. Most of the discussion on community indicators and their use has been in the context of environmental sustainability (Reed et al. 2005, 2006; Bossel 2001; Fraser et al. 2005). There is a need to explore these further in other contexts.

Accountability and learning

Accountability as a main purpose of evaluation may contradict other purposes, as it is generally achieved through independence and rigour, which can lead to tensions with approaches adopted to pursue additional purposes. Baker et al. (2007) highlight potential conflicts between accountability and learning, and

Patton (2008) stresses that evaluations required by donors often become an end in themselves, which can undercut utility for other stakeholders.

Van de Berg (2004) describes a similar dilemma between rigour and comprehensiveness, which can be caused by the limited understanding of the logical links between input-output-outcome in result-based evaluation frameworks (Rennekamp and Arnold 2009). Despite wide application of logic modelling in development interventions, application in extension did not happen until the mid-1990s (Taylor-Powell and Boyd 2008). If we are to learn from evaluations, they need to be designed in such a way that the information they generate helps to confirm the presumed linkages between actions and outcomes. Only then can the evaluation theory for extension systems be strengthened. Many of the extensions reviewed by Pound et al. (2011) fail to develop these linkages. It is important to develop more flexible and non-linear models which can incorporate various directions of behaviour change and identify the 'mediators' between interventions and behaviour changes and eventual outcomes (Braverman and Engle 2009). The approach also needs to be open to acknowledging a combination of factors such as commodity prices, marketing systems, level of education, policy context, appropriateness of innovation, etc. (Hoffmann et al. 2009).

There are trends in institutional practices aimed at separating accountability functions from wider evaluation purposes. Examples here are UK, DEC and Sida, the later having an Aid Watch Dog for accountability purposes and an internal evaluation department focusing on utilization-based evaluation with a strong learning orientation.

Learning/ capacity strengthening

Very few of the case studies reviewed by Pound et al. (2011) emphasize 'learning' and 'capacity building' as an important purpose of the evaluation process. This is also reflected in the fact that procedures and guidelines for sharing and disseminating evaluation findings are frequently lacking, which is an important weakness across different evaluation approaches. The identification of users and the planning for communicating findings to these users should be a key step in evaluation design. Capacity strengthening applies to the range of stakeholders and their organisations involved in the evaluation process. It can refer to improved capacities in evaluation management of external and internal evaluators, but it can also refer to opportunities for community empowerment (Fraser et al. 2005). Learning and capacity strengthening are closely linked to participation, user involvement and demand orientation as well as to accountability and learning, as discussed above.

Adaptive - change orientation

Evaluation approaches which are utilization-based and participatory are more likely to be able to adapt over time as community circumstances change (Pretty 1995; Carruthers and Tinning 2003; Freebairn and King 2003). Furthermore, due to their change orientation, they are more likely to build users' capacities to address future problems and lead to local action to deal with current problems. The limitations of more conventional methods to capture unpredictable but relevant effects have been criticized, as this leads to less adaptive interventions with limited contributions of the evaluation findings for programme improvement (Murray 2000).

Pluralistic delivery agents – partnerships/ networking

This is an aspect that has received little attention in evaluation literature and can perhaps best be described by the additional EC criterion of coherence/ complementarity. Emerging RAS systems consist of a range of potential delivery agents and this requires collaboration and effective allocation of tasks among them. In terms of evaluation approach, it requires a detailed stakeholder analysis and careful identification of potential participants in the evaluation process. It is therefore partly linked to the aspects discussed under the 'participation' section above. There are challenges to developing wider inclusion of service providers in RAS evaluations, particularly those who are not defined in terms of 'project partners'. There is still an important shift to be made in terms of thinking about RAS as a network which operates across government and NGO agencies and the private sector and outside of particular projects.

Gaps

As pointed out by various authors, there is a gap in evaluation studies which combine different approaches in order to capture the complexity of the RAS situation and to cater for the demands and requirements of different user groups. Findings from recent reviews of evaluations of Farmer Field Schools (FFS), which can be described as multi-dimensional and complicated/complex in terms of social and technical interactions, suggest the need for a combination of evaluation methods in order to evaluate impact. The Box below systematizes some of the key findings from recent studies:

Box 4: Lessons from evaluations of Farmer Field Schools (FFS)

A study of different evaluation methods for FFS conducted by Mancini and Jiggins (2008) suggested that opinions on the evaluation approach are divergent and will most likely remain divergent across stakeholders, and that there is a need for a range of evaluation approaches and methodological innovations to address evolving challenges posed by assessing complexities in development. They conclude that mono-disciplinary studies with pre-determined objectives are no longer considered sufficient to evaluate development interventions centred on people's empowerment. This is supported by van der Berg (2004), who reviewed 25 case studies of FFS evaluations and found out that all the studies were designed to be either statistically rigorous (but with a restricted scope) or comprehensive (but with limited coverage), but never both, which had negative consequences for their overall conclusions on FFS performance. Similarly, Njuki et al. (2008), based on a study conducted in Malawi, emphasize the need for 'hybrid' methodologies using elements of participatory and conventional approaches for evaluating research and development programmes.

A further area of intervention should focus on better targeted dissemination of evaluation findings to enhance learning. There is an urgent need to disseminate and share findings not only within projects/programmes but also with the wider extension evaluation community so that professional practice can be further informed.

Ways forward

The review of the literature suggests that to facilitate the identification and implementation of 'hybrid' approaches, there is a need to develop guidance for evaluation planning which helps in the selection of evaluation approaches appropriate for complex situations. An initial informal inquiry into the usefulness of an extension evaluation toolbox has revealed that there is more demand for a 'concept and principle' box, aimed at bringing different stakeholders' perspectives and evaluation demands closer together, rather than the development of an additional evaluation toolbox.

There are a large number of available guidelines and tools available for evaluation – some generic, like the EC evaluation guidelines, and some more specific. The specific tools and techniques used should be consistent with the principles underpinning the evaluation and its objectives and tailored to facilitate exploration of the evaluation questions that have been defined within the time and resources available.

Findings based on the literature have indicated a particular need for principles and guidance concerning:

- Decision-making on the level and extent of stakeholder participation in extension evaluation
- Development of flexible and non-linear programme theory models
- Designing mixed method evaluations which address both the impacts of what has been done and the strategic and institutional positioning of RAS interventions
- Procedures and guidelines for sharing and disseminating evaluation findings to different users
- Using the criteria of coherence and complementarity.
- Combining and sequencing different approaches and tools in evaluation

References

- ActionAid. 2009. *Who's really fighting hunger? Hunger Free ScoreCard* 2009. Johannesburg: ActionAid.
- AED/USAID n.d. Transforming the Kenyan Dairy Feeds System to Improve Farmer Productivity and Livelihoods. A SCALE case study, Kenya. AED Washington DC. www.aed.org/Publications/upload/Kenya_Case_revised_020608.pdf (accessed 19 July 2011)
- African Farm Radio Research Initiative 2009. The effect of participatory radio campaigns on agricultural improvement uptake. A preliminary case study assessment. Farm Radio International. www.farmradio.org/english/partners/afrri/casestudy-report.pdf(accessed 19 July 2011)
- Alkin, M. C., and C. A. Christie. 2004. An evaluation theory tree. In Evaluation Roots, ed. M. C. Alkin, 12–65. Thousand Oaks: Sage.
- Anderson, J. R. 2007. Agricultural advisory services. A background paper for "Innovating through science and technology", Chapter 7 of the World Development Report 2008. http://siteresources.worldbank.org/INTWDR2008/Resources/2795087-1191427986785/Anderson_AdvisoryServices.pdf (accessed 6 June 2011).
- Baker J., I. Christoplos, S. Dahlgren, S. Frueh, T. Kliest, Z. Ofir, and P. Sandison. 2007. Peer Review of the Evaluation Function at WFP. Stockholm: Sida.
- Bakewell, O., and A. Garbutt. 2005. The Use and Abuse of the Logical Framework Approach. Stockholm: Sida.

- Barker, B. S. 2007. Stakeholders' input on 4-H science and technology program area exploratory study. Journal of Extension [On-line], 45(2) Article 4RIB6. Available at: www.joe.org/joe/2007april/rb6.php (accessed 18 July 2011)
- Birner, R., K. Davis, J. Pender, E. Nkonya, P. Anandajayasekeram, J. Ekboir, A. Mbabu, D. Spielman, D. Horna, S. Benin, and M. Cohen. 2006. From 'best practice' to 'best fit': A framework for designing and analyzing pluralistic agricultural advisory services worldwide. DSGD Discussion Paper No. 37, EPTD Discussion Paper No. 155, FCND Discussion Paper No. 210, ISNAR Discussion Paper No.5. Washington: International Food Policy Research Institute. www.ifpri.org/sites/default/files/pubs/divs/dsgd/dp/papers/dsgdp37.pdf (accessed 6 June 2011).
- Bossel, H. 2001. Assessing viability and sustainability: a systems-based approach for deriving comprehensive indicator sets. *Conservation Ecology* 5, 12 (online).
- Braverman, M. T., and M. E. Arnold. 2008. An evaluator's balancing act: Making decisions about methodological rigor. In *Program evaluation in a complex organizational system: Lessons from Cooperative Extension,* ed. M. T. Braverman, M. Engle, M. E. Arnold, and R.A. Rennekamp, New Directions for Evaluation No. 120, 71–86. San Francisco: Jossey-Bass.
- Braverman, M. T., and M. Engle. 2009. Theory and rigor in Extension program evaluation planning. *Journal of Extension* 47(3). www.joe.org/joe/2009june/a1.php (accessed 6 June 2011).
- Carruthers, G., Tinning, G., 2003. Where, and how, do monitoring and sustainability indicators fit into environmental management systems? *Australian Journal of Experimental Agriculture* 43, 307–323.
- Chen, H. T. 1990. *Theory-driven evaluations.* Newbury Park, CA: Sage Publications
- Christoplos, I. 1996. *Poverty, Pluralism and Extension Practice.* Gatekeeper Series No. 64. International Institute for Environment and Development.
- Christoplos, I. 2010. Mobilizing the potential of rural and agricultural extension. Rome: Food and Agriculture Organization (FAO) of the United Nations and Global Forum for Rural Advisory Services (GFRAS).

- CIDA. 2004. *CIDA Evaluation Guide.* Quebec: Canadian International Development Agency (CIDA). www.acdi-cida.gc.ca/INET/IMAGES.NSF/vLUImages/Performancereview5/$file/English_E_guide.pdf (accessed 9 September 2010).
- CIDA. 2006. *Organisation Assessment Guide.* Quebec: Canadian International Development Agency (CIDA). www.acdi-cida.gc.ca/INET/IMAGES.NSF/vLUImages/Performancereview6/$file/OA%20Guide-E.pdf (accessed 9 September 2010).
- Commission of the European Communities. 2007. Responding to Strategic Needs: Reinforcing the use of evaluation. Communication to the Commission from Ms Grybauskait☐ in agreement with the President. Brussels, 21 February 2007. SEC(2007)213. http://ec.europa.eu/dgs/secretariat_general/evaluation/docs/eval_comm_sec_2007_213_en.pdf (accessed 6 June 2011).
- Cook, T. and Campbell, D.T. 1979. *Quasi-Experimentation: Design and analysis issues for field settings.* Boston, MA: Houghton Mifflin.
- DANIDA. 2006a. *Evaluation Policy 2006.* Copenhagen: DANIDA. www.um.dk/NR/rdonlyres/FFCDA746-F919-436E-B9AE-BADAE6BE8308/0/Danidasevaluationpolicy2006.pdf (accessed 9 September 2010).
- DANIDA 2006b. *Evaluation Guidelines 2006.* Copenhagen: DANIDA. www.um.dk/NR/rdonlyres/4BA486C7-994F-4C45-A084-085D42B0C70E/0/Guidelines2006.pdf (accessed 9 September 2010).
- Davies, R. J., and J. Dart. 2005. The most significant change (MSC) technique: A guide to its use. www.mande.co.uk/docs/MSCGuide.pdf (accessed 6 June 2011).
- Deshler, D. 1997. Evaluating extension programmes. In *Improving Agricultural Extension. A Reference Manual.* Rome: FAO. www.fao.org/docrep/w5830e/w5830e0d.htm#TopOfPage (accessed 6 June 2011).
- Donaldson, S. I., M. Q. Patton, D. M. Fetterman, and M. Scriven. 2010. The 2009 Claremont Debates: The promise and pitfalls of utilization-focused and empowerment evaluation. *Journal of MultiDisciplinary Evaluation* 6(13):15–57.
- Duniform, R., M. Duttweiler, K. Pillemer, D. Tobias, and W. M. K. Trochim. 2004. Evidence-based Extension. *Journal of Extension* 42(2):2FEA2. www.joe.org/joe/2004april/a2.php (accessed 6 June 2011).

- Duignan, P. 2008. Reframing program evaluation as part of strategic information collection for sector decision-making. Outcomes Theory Knowledge Base Article No. 221. http://knol.google.com/k/reframing-program-evaluation-as-part-of-collecting-strategic-information-for# (accessed 19 July 2011)
- Duignan, P. 2009a. Evaluation types: Formative/developmental, process, impact/outcome evaluation. Outcomes Theory Knowledge Base Article No. 256. http://knol.google.com/k/paul-duignan-phd/-/2m7zd68aaz774/119 (accessed 6 June 2011).
- Duignan, P. 2009b. Implications of an exclusive focus on impact evaluation in 'what works' evidence-based practice systems. Outcomes Theory Knowledge Base Article No. 223. http://knol.google.com/k/paul-duignan-phd/implications-of-an-exclusive-focus-on/2m7zd68aaz774/46 (accessed 6 June 2011).
- Duignan, P. 2009c. Selecting impact/outcome evaluation designs: a decision-making table and checklist approach. Outcomes Theory Knowledge Base Article No. 256. http://knol.google.com/k/paul-duignan-phd/selecting-impactoutcome-evaluation/2m7zd68aaz774/115 (accessed 6 June 2011).
- Duignan, P. 2009d. Terminology in evaluation: Approaches, types (purposes), methods, analysis techniques and designs. Outcomes Theory Knowledge Base Article No. 259. http://knol.google.com/k/paul-duignan-phd/terminology-in-evaluation-approaches/2m7zd68aaz774/120 (accessed 6 June 2011).
- Earl, S., F. Carden, and T. Smutylo. 2001. Outcome Mapping: Building Learning and Reflection into Development Programs. Ottawa: IDRC. www.idrc.ca/en/ev-9330-201-1-DO_TOPIC.html (accessed 6 June 2011).
- European Commission. 2004. *Evaluating EU Activities – A Practical Guide for the Commission Services.* Brussels: European Commission http://ec.europa.eu/dgs/secretariat_general/evaluation/docs/eval_activities_en.pdf (accessed 9 September 2010).
- European Commission. n.d. *Evaluation Standards.* http://ec.europa.eu/dgs/secretariat_general/evaluation/docs/standards_c_2002_5267_final_en.pdf (accessed 9 September 2010).

- EuropeAid 2006. Evaluation Methods. http://ec.europa.eu/europeaid/evaluation/methodology/methods/mth_pro_en.htm (accessed 14 July 2011)
- FAO. 2007. Evaluation in FAO: Institutional Arrangements, Policies and Methods. Rome: Food and Agriculture Organization (FAO). www.fao.org/eims/pbe/eims_search/1_dett.asp?calling=simple_s_result&publication=&webpage=&photo=&press=&lang=en&pub_id=235302 (accessed 9 September 2010).
- FAO. 2009. Programme Evaluation Report. Rome: Food and Agriculture Organization (FAO). ftp://ftp.fao.org/docrep/fao/meeting/018/k6197e.pdf (accessed 9 September 2010).
- Fetterman, D.M., Kaftarian, S.J., & Wandersman, A. (Eds.). 1996. *Empowerment evaluation: Knowledge and tools for self-assessment and accountability.* Thousand Oaks, CA: Sage.
- Fetterman, D. M., and A. Wandersman. 2004. *Empowerment Evaluation Principles in Practice.* New York: Guilford Publications.
- Fraser, E., Mabee, W., Figge, F., 2005. A framework for assessing the vulnerability of food systems to future shocks. *Futures* 37, 465–479.
- Freebairn, D.M., King, C.A., 2003. Reflections on collectively working toward sustainability: indicators for indicators! *Australian Journal of Experimental Agriculture* 43, 223–238.
- Funnell, S. C., and P. J. Rogers. 2011. *Purposeful Program Theory: Effective Use of Theories of Change and Logic Models.* San Francisco: Jossey-Bass.
- Gasper, D. 2000. Evaluating the 'logical framework approach' towards learning-oriented development evaluation. *Public Administration and Development* 20(1):17–28.
- Global Donor Platform for Rural Development. 2006. *On Common Ground: A Joint Donor Concept on Rural Development.* Bonn: Global Donor Platform for Rural Development.
- Global Donor Platform for Rural Development. 2008. Agricultural Sector Experiences in Implementing the Paris Declaration on Aid Effectiveness. Bonn: Global Donor Platform for Rural Development.

- GFRAS. 2010a. Five Key Areas for Mobilising the Potential of Rural Advisory Services. GFRAS Brief 1, October 2010. Global Forum for Rural Advisory Services (GFRAS), c/o Agridea, Eschikon 28, 8315 Lindau, Switzerland. www.g-fras.org/en/knowledge/gfras-publications# (accessed 19 July 2011)
- GFRAS. 2010b. Providing space for advocacy and leadership on rural advisory services Strategy and Work Plan for Inception Phase (January 2010 – June 2011). Global Forum for Rural Advisory Services (GFRAS), c/o Agridea, Eschikon 28, 8315 Lindau, Switzerland. www.g-fras.org/fileadmin/UserFiles/GFRAS-documents/GFRAS-Strategy-for-Inception-Phase_Jan2010-June2011.pdf (accessed 19 July 2011)
- Greene, J. 1988. Stakeholder participant and utilization in program evaluation. *Evaluation Review,* 12: 91–116.
- GTZ. 2006. Working for sustainable results. Evaluation at GTZ. Eschborn: GTZ. www2.gtz.de/dokumente/bib/06-0796.pdf (accessed 9 September 2010).
- GTZ. 2007. *Results Monitoring 2007: Evaluation Report on the Work of GTZ and Its Partners. 10th Cross-section Analysis 2006–2007.* Eschborn: GTZ. www.gtz.de/en/dokumente/Evaluation-Report-0801.pdf (accessed 9 September 2010).
- Guba, E.G. and Lincoln, Y.S. 1989. *Fourth generation evaluation.* Save, Newbury Park, CA.
- Guion, L. A. 2002. Triangulation: Establishing the Validity of Qualitative Studies. Department of Family, Youth and Community Sciences, Florida Cooperative Extension Service, FCS6014. Gainesville: Institute of Food and Agricultural Sciences, University of Florida.
- Guijt, I. 2004. ALPS in Action: A Review of the Shift in ActionAid towards a new Accountability, Learning and Planning System. Action Aid International. www.actionaid.org.uk/_content/documents/ALPSReview.pdf (accessed 14 July 2011)
- Hall, A., V. Rasheed Sulaiman, N. Clark, and B. Yoganand. 2003. From measuring impact to learning institutional lessons: An innovation systems perspective on improving the management of international agricultural research. *Agricultural Systems* 78(2):213–241.
- Hoffmann, V., M. Gerster-Bentaya, A. Christinck, and M. Lemma, eds. 2009. *Basic Issues and Concepts.* Vol. 1 of *Rural Extension.* Weikersheim: Margraf Publishers.

- IAASTD. 2009. *Agriculture at a Crossroads: Synthesis Report.* Washington: Island Press.
- IDRC. 2007. *IDRC's Approach to Evaluation.* Ottawa: International Development Research Centre (IDRC). www.idrc.ca/uploads/user-S/12194973371IDRC's_Approach_to_Evaluation.pdf (accessed 9 September 2010).
- IDRC. n.d. *Guiding Principles of IDRC's Evaluation Unit.* Ottawa: International Development Research Centre (IDRC). www.idrc.ca/uploads/user-S/12095810441Evaluation_Unit_Guiding_Principles.pdf (accessed 21 September 2010).
- IFAD. 2003. *IFAD Evaluation Policy.* Rome: International Fund for Agricultural Development (IFAD). www.ifad.org/pub/policy/oe.pdf (accessed 9 September 2010).
- IFAD. 2009. *Evaluation Manual: Methodology and Processes.* Rome: International Fund for Agricultural Development (IFAD). www.ifad.org/evaluation/process_methodology/doc/manual.pdf (accessed 9 September 2010).
- Jones, H. and Mendizabal, E. 2010. 'Strengthening learning from research and evaluation: going with the grain', report for IACDI. London: ODI.
- Lawrenz, F. & Huffman, D. 2003. How can multi-site evaluations be participatory? *American Journal of Evaluation,* 24(4), 331–338.
- Leeuwis, C., and A. van den Ban. 2004. *Communication for Rural Innovation: Rethinking Agricultural Extension.* 3rd ed. Oxford: Blackwell Publishing.
- Ljungman, Cecilia M., Helge Rønning, Tejeshwar Singh, Henrik Steen Pedersen et al. 2005. *Sida's Work with Culture and Media.* Main Report. Sida Evaluation 04/38.
- Mancini, F. and Jiggins, J. 2008. Appraisal of methods to evaluate farmer field schools', *Development in Practice,* 18:4,539–550.
- Murray, P 2000. Evaluating participatory extension programs: challenges and problems. *Australian Journal of Experimental Agriculture* 2000 Vol. 40 No. 4 pp. 519–526
- Nagel, U. J. 2003. Back on the Agenda: Extension and Its Institutional Linkages. Some Personal Observations on the Re-discovery of a Key Player. Paper presented at the Tropentag in Goettingen, Germany, 2003.

- Narayan, Deepa. 1993. Participatory Evaluation: Tools for Managing Change in Water and Sanitation. World Bank Technical Paper 207
- Neuchâtel Group. 1999. Common Framework on Agricultural Extension. Lindau: Neuchâtel Group. www.g-fras.org/fileadmin/UserFiles/Documents/Frames-and-guidelines/New-paradigms/Common-Framework-on-Agricultural-Extension.pdf (accessed 6 June 2011).
- Neuchâtel Group. 2000. Guide for Monitoring, Evaluation and Joint Analyses of Pluralistic Extension Support. Lindau: Neuchâtel Group. www.g-fras.org/fileadmin/UserFiles/Documents/Frames-and-guidelines/M_E/Guide-for-Monitoring-Evaluation-and-Joint-Analysis.pdf (accessed 6 June 2011).
- Neuchâtel Group. 2006. Demand-Driven Agricultural Advisory Services. Lindau: Neuchâtel Group. www.g-fras.org/fileadmin/UserFiles/Documents/Frames-and-guidelines/New-paradigms/Demand-Driven-Agricultural-Advisory-Services.pdf (accessed 6 June 2011).
- Njuki, J., M. Mapila, S. Kaaria, and T. Magombo. 2008. Using community indicators for evaluating research and development programmes: Experiences from Malawi. *Development in Practice* 18(4):633–642.
- NORAD. 2006. Evaluation Policy 2006–2010. Part 1 Strategic priorities, Part 2 Evaluation Programme 2006–2008, Part 3 Guidelines for Evaluation of Norwegian Development Cooperation. Oslo: NORAD.
- OECD. 1991. *DAC Principles for Evaluation of Development Assistance.* Paris: Development Assistance Committee.
- OECD. 1998. *Review of the DAC Principles for Evaluation of Development Assistance.* Paris: DAC Working Party on Aid Evaluation. www.oecd.org/dataoecd/63/50/2065863.pdf (accessed 6 June 2011).
- OECD. 2005/2008. The Paris Declaration on Aid Effectiveness and the Accra Agenda for Action. Paris: OECD. www.oecd.org/dataoecd/30/63/43911948.pdf (accessed 6 June 2011).

- OECD/DAC Working Party on Aid Effectiveness. 2008. *Aid Effectiveness: A Progress Report on Implementing the Paris Declaration. 3rd High level Forum on Aid Effectiveness, September 2–4, 2008, Accra, Ghana.* http://siteresources.worldbank.org/ACCRAEXT/Resources/Progress_Report-Full-EN.pdf (accessed 6 June 2011).
- OECD. 2010a. Evaluation in Development Agencies, Better Aid. Paris: OECD Publishing. http://dx.doi.org/10.1787/9789264094857-en (accessed 6 June 2011).
- OECD. 2010b. *Quality Standards for Development Evaluation.* DAC Guideline and Reference Series. www.oecd.org/dataoecd/55/0/44798177.pdf (accessed 6 June 2011).
- Patton, M. Q. 2002a. *Qualitative Research and Evaluation Methods.* Thousand Oaks: Sage Publications.
- Patton, M. Q. 2002b. Utilization-Focused Evaluation (U-FE) CHECKLIST. www.fmfi.org.za/wiki/index.php/Utilisation_Focused_Evaluation (accessed 6 June 2011).
- Patton, M.Q. 2007. Process use as a usefulism. In J. B. Cousins (Ed.), Process use in theory, research, and practice. *New Directions for Evaluation,* 116, 99–112.
- Patton, M. Q. 2008. *Utilization-Focused Evaluation.* Thousand Oaks: Sage Publications.
- Patton, M.Q. 2010. *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use.* New York, NY: Guilford Press.
- Pound, B., S. Gündel, A. Martin, and E. Apenteng. 2011. *Meta-evaluation of Extension Evaluation Case Studies.* Lindau: NRI/GFRAS.
- Preskill, H., and A. T. Coghlan, eds. 2003. *Using Appreciative Inquiry in Evaluation.* New Directions for Evaluation No. 100. San Francisco: Jossey-Bass.
- Pretty, J.N., 1995. Participatory learning for sustainable agriculture. *World Development* 23, 1247–1263.
- Rajalahti, R., J. Woelcke, and E. Pehu. 2005. *Monitoring and Evaluation for World Bank Agricultural Research and Extension Projects: A Good Practice Note.* Agriculture and Rural Development Discussion Paper 20. Washington: The International Bank for Reconstruction and Development / The World Bank. http://siteresources.worldbank.org/INTARD/Resources/ARD_DP20.pdf (accessed 14 July 2011)

- Reed, M.S., Fraser, E.D.G., Morse, S., Dougill, A.J., 2005. Integrating methods for developing sustainability indicators that can facilitate learning and action. *Ecology and Society* 10 (1): r3 (online).
- Reed, M.S; Fraser, E.D.G.; Dougill, A.J. 2006 An adaptive learning process for developing and applying sustainability indicators with local communities. *Ecological Economics* 59 (406–418) (online)
- Rennekamp, R. A., and M. E. Arnold. 2009. What progress, program evaluation? Reflections on a quarter-century of extension evaluation practice. *Journal of Extension* 47(3):3COM1.
- Reuber, M., and O. Haas. 2009. *Evaluations at the Deutsche Gesellschaft für Technische Zusammenarbeit* (GTZ GmbH) German Technical Cooperation. Eschborn: GTZ. www.gtz.de/en/dokumente/en-evaluations-at-gtz-reuber-haas-2009.pdf (accessed 9 September 2010).
- Scriven, M. 1972. Pros and cons about goal-free evaluation.: Evaluation comment. *The Journal of Educational Evaluation,* 3(4): 1-7.
- Sida. 2004. *Looking Back, Moving Forward. Sida Evaluation Manual.* Stockholm: Swedish International Development Cooperation Agency (Sida). www.alnap.org/pool/files/evaluation_manual_sida.pdf (accessed 13 July 2011)
- Stufflebeam, D. L. 2000. The Methodology of Metaevaluation as Reflected in Metaevaluations by the Western Michigan University Evaluation Center. *Journal of Personnel Evaluation in Education,* 14(1):95–125.
- Swanson, B. E., and R. Rajalahti. 2010. S*trengthening Agricultural Extension and Advisory Systems: Procedures for Assessing, Transforming, and Evaluating Extension Systems.* Agriculture and Rural Development Discussion Paper 45. Washington: World Bank.
- Taylor-Powell, E., and H. H. Boyd. 2008. Evaluation capacity building in complex organizations. In P*rogram Evaluation on a Complex Organizational System: Lessons from Cooperative Extension,* ed. M. T. Braverman, M. Engle, M. E. Arnold, and R. A. Rennekamp, New Directions for Evaluation No. 120, 55–69. San Francisco: Jossey-Bass.
- UNEG. 2005. Norms for Evaluation in the UN system. New York: United Nations Evaluation Group (UNEG).

- UNDP. 2009. *Handbook on Planning, Monitoring and Evaluating for Development Results.* New York: UNDP. www.undp.org/evaluation/handbook/documents/english/pme-handbook.pdf (accessed 9 September 2010).
- van den Berg, H. 2004. *IPM FFS: A synthesis of 25 impact evaluations. Report prepared for the Global IPM Facility.* Wageningen: The Netherlands.
- Weiss, C. H. 1995. Nothing as Practical as Good Theory: Exploring Theory-Based Evaluation for Comprehensive Community Initiatives for Children and Families", in J. Connell, A. Kubisch, L. B. Schorr, and C. H. Weiss (eds) New Approaches to Evaluating Community Initiatives, New York: Aspen Institute.
- White, H. 2009a. *Some Reflections on Current Debates in Impact Evaluation.* International Initiative for Impact Evaluation, Working paper 1. New Delhi: International Initiative for Impact Evaluation (3ie). www.3ieimpact.org/admin/pdfs_papers/11.pdf (accessed 6 June 2011).
- White, H. 2009b. *Theory-Based Impact Evaluation: Principles and Practice.* International Initiative for Impact Evaluation, Working paper 3. New Delhi: International Initiative for Impact Evaluation (3ie). www.3ieimpact.org/admin/pdfs_papers/51.pdf (accessed 6 June 2011).
- W. K. Kellogg Foundation. 2004a. Evaluation Handbook. www.wkkf.org/knowledge-center/resources/2010/W-K-Kellogg-Foundation-Evaluation-Handbook.aspx (accessed 6 June 2011).
- W. K. Kellogg Foundation. 2004b. Logic Model Development Guide. Using Logic Models to Bring Together Planning, Evaluation, and Action. www.wkkf.org/knowledge-center/resources/2006/02/WK-Kellogg-Foundation-Logic-Model-Development-Guide.aspx (accessed 6 June 2011).
- World Bank. 2004. Monitoring and Evaluation. Some tools, methods and approaches. Washington: The World Bank.
- World Bank. 2006. *Enhancing Agricultural Innovation: How to Go Beyond the Strengthening of Research Systems.* Washington: The International Bank for Reconstruction and Development / The World Bank. http://siteresources.worldbank.org/INTARD/Resources/Enhancing_Ag_Innovation.pdf (accessed 6 June 2011).

- World Bank. 2008. *World Development Report 2008: Agriculture for Development.* Washington: The International Bank for Reconstruction and Development / The World Bank.
- World Bank Group. n.d. Evaluation for better results. IEG Brochure. Washington: Independent Evaluation Group (IEG). http://siteresources.worldbank.org/EXTOED/Resources/ieg_brochure.pdf

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

The Global Forum for Rural Advisory Services (GFRAS) has commissioned work to develop guidance for the evaluation of extension (projects, programmes, tools and initiatives). The purpose of the overall project is to identify methods for better evaluation of extension through the development of a Guidance Kit for extension evaluation.

This review of literature on evaluation methods, in combination with a metaevaluation[1] of extension evaluation[2] case studies, is intended to be a resource for developing the Guidance Kit. It is envisaged that this paper will be of interest to those involved in managing and implementing evaluations of rural advisory services as well as to extension and evaluation specialists.