Available online at www.sciencedirect.com
SCIENCE DIRECT
Agricultural Systems 78 (2003) 243–265
AGRICULTURAL SYSTEMS
www.elsevier.com/locate/agsy

Impact pathway evaluation: an approach for achieving and attributing impact in complex systems

Boru Douthwaite[a,*], Thomas Kuby[b], Elske van de Fliert[c], Steffen Schulz[d]

[a]*Centro Internacional de Agricultura Tropical (CIAT), Cali, Colombia*
[b]*Deutsche Gesellschaft für Technische Zusammenarbeit GmbH (GTZ), Eschborn, Germany*
[c]*Food and Agriculture Organization (FAO) Regional Vegetable Program, Hanoi, Viet Nam*
[d]*Swiss Organisation for Development and Cooperation (Intercooperation), Katmandu, Nepal, India*

Abstract

Agricultural development is fundamentally a social process in which people construct solutions to their problems, often by modifying both new technologies and their own production systems to take advantage of new opportunities offered by the technologies. Hence, agricultural change is an immensely complex process, with a high degree of non-linearity. However, current 'best practice' economic evaluation methods commonly used in the CGIAR system ignore complexity. In this paper we develop a two-stage monitoring, evaluation and impact assessment approach called impact pathway evaluation. This approach is based on program-theory evaluation from the field of evaluation, and the experience of the German development organization GTZ (Deutsche Gesellschaft für Technische Zusammenarbeit GmbH). In the first stage of this approach, a research project develops an impact pathway for itself, which is an explicit theory or model of how the project sees itself achieving impact. The project then uses the impact pathway to guide project management in complex environments. The impact pathway may evolve, based on learning over time. The second stage is an ex post impact assessment sometime after the project has finished, in which the project's wider benefits are independently assessed. The evaluator seeks to establish plausible links between the project outputs and developmental changes, such as poverty alleviation. We illustrate the usefulness of impact pathway evaluation through examples from Nigeria and Indonesia.

Keywords: Impact assessment; Participatory approaches; Program theory; Integrated natural resource management; Integrated pest management; *Striga*; Nigeria; Indonesia

* Corresponding author. Tel.: +37-353-98-25313; fax: +37-353-98-25313.
E-mail address: b.douthwaite@cgiar.org, boru@douthwaite.net (B. Douthwaite).

0308-521X/$ - see front matter
doi:10.1016/S0308-521X(03)00128-8

1. Introduction

The way people behave, including scientists, is determined to a large extent by their belief structures formed during their education. In science another word for 'belief structures' is paradigm, which is defined as: "a set of assumptions, concepts, values, and practices that constitutes a way of viewing reality for the community that shares them" (Houghton Mifflin, 2000).

Professor Niels Röling of Wageningen University believes that the paradigms that people are operating from should be made explicit rather than implicit or tacit (from the Foreword of Douthwaite, 2002). One reason for this, as Thomas Kuhn (1970) points out in his highly influential book *The Structure of Scientific Revolutions*, is that research is not about discovering the unknown, but rather "a strenuous and devoted attempt to force nature into the conceptual boxes supplied by professional education". This creates what Rogers (1995) describes as 'invisible colleges' of researchers who have similar educational backgrounds, quote each other's work, use each other's methods while remaining largely unaware of what lies outside their 'college'.

The dominant paradigm within the 16 research centres that constitute the Consultative Group on International Agricultural Research (CGIAR) is positivism (Douthwaite et al. 2001a). The CGIAR system is described by Horton and McKay in this issue. This set of belief structures sees innovation as a rather simple and linear process. Positivism has successfully underpinned the CGIAR's early successes in breeding high yielding rice and wheat varieties that helped catalyze the Green Revolution, and much of the CGIAR's other work. As a result impact assessment in the CGIAR System now largely takes place within an 'invisible college' with positivism as the dominant paradigm, and agricultural economics as the dominant discipline. However, emergent understanding of rural development as a social and complex process is leading to positivism giving way to constructivism. Within the CGIAR the constructivist movement is largely housed within Integrated Natural Resource Management (INRM). INRM is defined as "an approach to research that aims at improving livelihoods, agro ecosystem resilience, agricultural productivity and environmental services" (Anon 2002). As such INRM includes Integrated Pest Management which has been one of the pioneers of a constructivist outlook in the CGIAR System. In this paper we show that INRM, and constructivist-based research in general, require different types of evaluation methods to complement existing economic impact assessment methods. These approaches can be found outside agriculture in the field of evaluation.

2. Why shifting research paradigm changes evaluation requirements

Positivism is associated with "hard" science, that is, science that sets up hypotheses and tests them with repeatable and quantifiable experiments. 'Hard' scientists (e.g., most natural scientists and economists) are trained to believe that the world they experience has an independent reality which they are discovering in their experiments. From the repeatability principle follows the idea that knowledge gained in this way is independent from context and separate from the knower, hence technologies built on scientific principles will work independent from the people who use them. From this follows the idea that technology that works under a certain set of agro-ecological and economic conditions can be transferred to a similar area, so long as the technology hardware (its physical manifestation) and software (instructions on how to replicate and use it) are faithfully reproduced. The social characteristics of the people adopting, and the way the technology is introduced, do not really matter.

The paradigm that underpins much of INRM, IPM and participatory research is constructivism. While positivists consider knowledge as being independent of context and therefore to be passively received 'as is' and 'mapped on' to a learner's brain, constructivists see the learning process as an active one where the learner 'constructs' knowledge by fitting new information into his or her existing ways of seeing the world. This construction process is social because part of understanding new phenomena is undertaken as a group through negotiation. Technical innovation—the process of making a new technology work—is a learning process (Rosenberg, 1982) and hence is also a social process (Bijker and Pinch, 1988; Röling, 1996). While positivists see farmers as being essentially passive recipients of extension messages, constructivists see the role of an extension worker much more as a facilitator fostering a 'social construction' process. This 'social construction' involves farmers experimenting with the new technology in the process of making sense of it and adapting it to their own socio-economic, cultural and agro-ecological conditions. Hence, constructivists see research continuing well into the extension phase, and this has important implications for the monitoring and evaluation of research.

A second area of difference that is relevant to evaluation is that constructivists see the legitimisation of new knowledge as being tied to its use through what Lyotard (1984) calls performativity—how well it serves peoples' purposes in the real world. Positivists, on the other hand, see publication in a peer-reviewed journal as the main validation of the quality of knowledge.

Fig. 1 attempts to classify important agricultural research traditions according to their dominant paradigm and whether the focus is on a local or general scale. The figure also shows the evolution in thinking that has taken place in international agricultural research, starting with the Green Revolution in the 1960s and 1970s. The Green Revolution is placed in the Positivist-General cell because the plant breeding that underpinned it was predominantly carried out on-station with little reference to farmers' local wishes and requirements. For example, early IRRI high yielding varieties were not bred for local taste preferences.

The Farming Systems Research (FSR) movement began in the early 1970s by field practitioners who saw that in developing countries "technologies recommended as a result of agricultural research investments were, in general, inappropriate to the priorities and circumstances of small farmers" (Collinson, 2000, p. 2). FSR's founders saw that the failure of many well intentioned agricultural programs were due to the fact that researchers failed to properly understand the realities and priorities of the intended beneficiaries—small-scale farmers, particularly those in resource poor areas. Early FSR is placed in the Positivist-Local cell to reflect that the early originators and proponents of FSR were mainly natural resource scientists and agricultural economists with a positivist grounding. As a result early FSR resembled market research (Röling, 1988)—it used tools to extract a better understanding of farmers' constraints which researchers then used to fine-tune their products. This resulted in a shortfall that McCown (2001, p. 8) describes as ''relocation-to-the-farm not equating to becoming-situated-in-the-farm''. The result was technologies that were moulded by scientists' perceptions of farmers' realities but not directly by farmers themselves.

The development of participatory approaches, often by FSR practitioners, was partly in response to this problem. Participatory approaches attempted to enlist farmers as active co-developers of new technology rather than seeing farmers as more passive providers of information and evaluation information, as early FSR had done. The constructivist paradigm provides a natural basis for participatory approaches.

INRM attempts to blend together both 'hard' and 'soft' science and thus are shown in Fig. 1 spanning both the positivist and constructivist domains. INRM contains within it reductionist research to understand basic biological processes and develop prototype technologies. On the other hand INRM is firmly grounded in the constructivist paradigm because it recognises that, amongst other things, dissemination efforts should focus on replicating the social and organizational *processes* involved in bringing about technological change, rather than concentrating on the transfer of *technology* (Sayer and Campbell, 2001).

Fig. 1 also shows INRM spanning the Local and General domains. This reflects the fact that publicly-funded international research needs to be producing international public goods, that is knowledge and technologies that have applicability

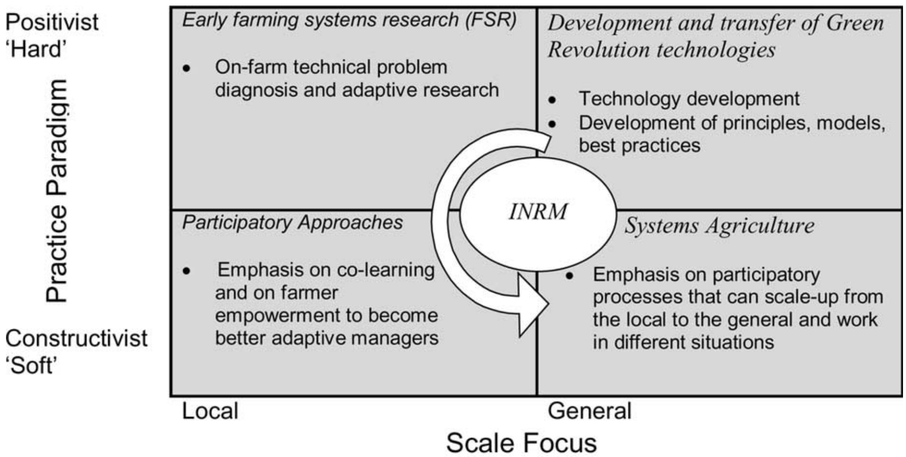

Fig. 1. The paradigms in international agricultural research and development (adapted from McCown, 2001).

beyond the local. INRM can 'act local' but they must also 'think global'. This means thinking about going to scale.

The concept of scaling is crucial to INRM. However the term has several meanings. In this paper we distinguish between three types, although discuss only the first two:

1. Scaling-out—innovation spread from farmer to farmer, community to community, within the same stakeholder groups;
2. Scaling-up—an institutional expansion from grassroots organizations to policy makers, donors, development institutions, and other stakeholders key to building an enabling environment for change.
3. Spatial scaling-up—the widening of scale of operation from, for example, experimental plot, to field, to farm, to watershed, etc.

Scaling-out and scaling-up processes are illustrated graphically in Fig. 2. Both are linked because as a change spreads further geographically the greater the chances of influencing those at higher levels, and likewise, as one goes to higher institutional levels then the greater the chances for horizontal spread. Fig. 2 also shows that interventions at a higher scale can affect scaling up processes at lower ones.

Embedded in the concepts of scaling up and out is the idea that technological change is brought about by the formation and actions of networks of stakeholders

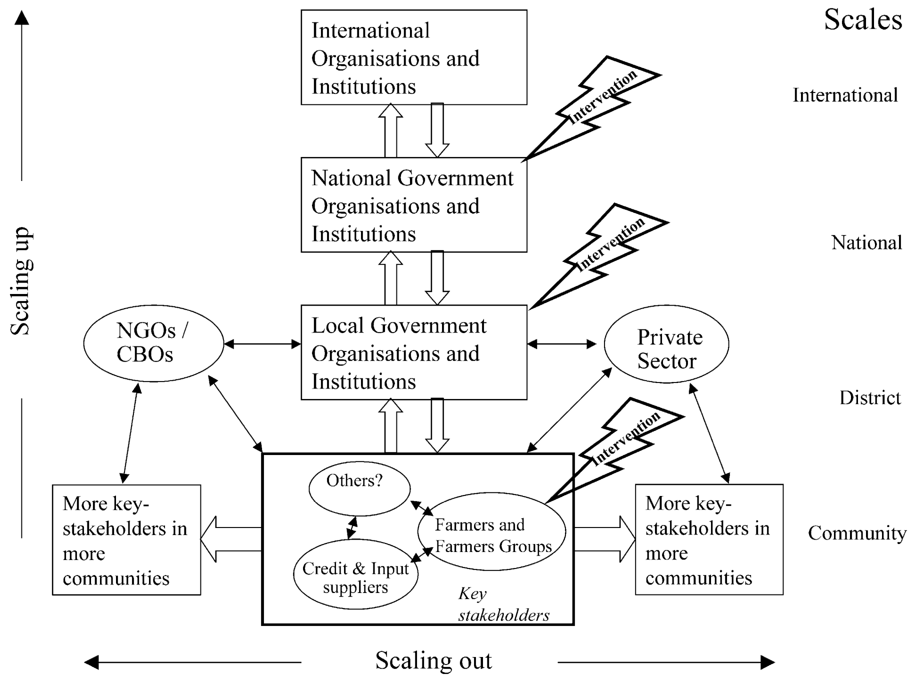

Fig. 2. Concepts of scaling-out and scaling-up (adapted from IIRR, 2000).

in what is essentially a social process of communication and negotiation. This is an important departure from positivist 'Green Revolution'.

This departure has profound implications for evaluation of INRM, and constructivist-based research in general. In the past the main outputs of the CGIAR system have been technologies, largely improved germplasm, not processes. The high yielding rice and wheat varieties that spawned the Green Revolution gave such large yield increases in relatively simple irrigated systems, when used together with some inorganic inputs, that the technology package virtually sold itself (Douthwaite et al., 2001a). This made transfer of technology very easy and meant that CGIAR scientists did not have to worry about extension, or scaling out and up. Instead they concentrated on the 'upstream' breeding of higher yielding varieties with more pest and disease resistance and left their collaborators in the National Agricultural Research and Extension Systems (NARES) to look after the rest.

As technologies go, high yielding varieties introduced into irrigated systems, was a special case. Farmers already knew how to grow crops, how to save seed, and were able to control their production environments to benefit from the improved varieties. Other technologies, in particular NRM and pest management ones, as well as improved germplasm introduced into poorer and more complex rainfed systems, require farmers to learn much more, and are not so easily replicated and promulgated (Douthwaite et al., 2001a). Indeed, the evolution of thinking shown in Fig. 1 has been driven by the attempt to achieve reasonable adoption of more complex technologies in more complex systems. It has led to the realisation that solutions to complex problems cannot be solved on-station but need to be built up in situ in farmers' fields, taking full advantage of farmer's knowledge and innovative abilities. These location-specific solutions, however, are not international public goods unless they can be adapted and applied more generally. The international public goods are therefore the location-specific solutions together with the methods to identify where they may also work, and the cost-effective and generally applicable *processes* to introduce them into these new areas.

As a result of the Green Revolution and the dominance of positivist-trained scientists in the CGIAR system, evaluation has focused on the economic impact assessment of technologies, largely to assist in resource allocation decisions and to show accountability to donors. Best practice in this 'hard' type of evaluation is represented by the Alston et al. (1995) book *Science under Scarcity*. The book only describes economic evaluation procedures and the authors dismiss other evaluation and impact assessment methods by saying that they:

> rarely attempt to establish a systematic causal relationship between the costs and benefits of research, and as a result, they are most unlikely to yield any meaningful indications of the economic effects of research. ... Therefore, they are not useful for informing allocation decisions. (p. 501–502)

In two sentences the authors dismiss out of hand a whole spectrum of evaluation approaches developed in the field of evaluation to assess social programs and to make resource allocation decisions. However, it is exactly these approaches that are likely to be relevant to INRM, given that its international public goods are going to include social processes.

3. Learning evaluation from others

The idea that impact assessment in the CGIAR system might have much to learn from the field of evaluation is not new. Horton (1998) asks: "Why is cost-benefit practically the only systematic method used? And why have broader program theory and expertise been virtually ignored?" His answer agrees with our analysis. "The answers appear to stem fundamentally from the 'hard science' culture of agricultural research organisations."

In May 2000 the Standing Panel on Impact Assessment (SPIA) of the Technical Advisory Committee (TAC) of the CGIAR system held a workshop on 'The Future of Impact Assessment in the CGIAR: Needs, Constraints and Options'. One of the invited speakers was Frans Leeuw, President of the European Evaluation Society (2000–2002), who presented a paper called "Program Evaluation and Social and Institutional Impact Assessment" in which he described program theory analysis (Leeuw, 2002).

In this paper we take up Horton and Leeuw's invitation to attempt to apply program theory evaluation to CGIAR research in complex settings. We begin first though by describing an existing conceptual framework for impact assessment, developed by GTZ (Deutsche Gesellschaft für Technische Zusammenarbeit GmbH), in which program theory evaluation can work.

3.1. The GTZ evaluation model

GTZ went through a major change in evaluation that began in 1999, in part through the efforts of one of the co-authors. The German Government, often through GTZ, is a major donor to the CGIAR system. GTZ also runs its own projects. Therefore, current thinking and 'best practice' evaluation within GTZ could provide an invaluable conceptual framework to guide how we might conduct evaluation of research carried out in complex settings.

The rationale GTZ gives for its new approach to evaluation is the growing body of knowledge that shows that linear, i.e., positivist, models of the innovation process GTZ used previously are over-simplistic and do not take into account the fact that innovation is a social process in which users 'socially construct' new technology (Kuby, 1999). In this rationale we see clear evidence of the relevance of GTZ's new approach to INRM and constructivist evaluation in general. Once one accepts that users are modifying technologies, and their own systems to technologies, and these adaptations affect adoption rates and who benefits and loses, then any innovation model becomes complex with high degrees of non-linearity. This makes economic impact assessment very difficult because it becomes virtually impossible to link project outputs with highly aggregated benefits like poverty eradication or food security. GTZ's new position is thus at odds with Alston et al.'s (1995) 'best practice' that relies on establishing a systematic causal relationship between the costs and benefits of research.

Fig. 3 shows the impact model that GTZ has developed to bring more plausibility and practicality to evaluation and impact assessment. The model splits evaluation/impact assessment into two parts. In the first stage individual projects carry out their own monitoring and evaluation to the point of assessing the impact of the direct benefits of the project. Assigning this role to the project comes from years of project experience in GTZ that has shown that: "as a rule, self-evaluation is more critical and better value for money than external monitoring—and that it makes a much greater contribution to learning, both in the projects and in the whole organization" (Kuby, 2000: p. 4).

The GTZ model then shows an 'attribution' gap between the project's direct benefits and its developmental outcomes, and it is exactly this gap that GTZ argue cannot be plausibly spanned using a linear, causal bridge. Instead GTZ envisages a second ex post impact assessment, carried out sometime after the project has finished, which would involve an impact assessor building a 'plausible' bridge between the project's direct benefits and wider level impacts. This would consist of a persuasive case being argued about how the direct benefits of the project have contributed to development progress. To be persuasive the case would use triangulation of data sources as far as possible, and blend together quantitative analysis with qualitative data and verbal testimony. Case study methodology, as described by GAO (1987), Yin (1994) and others, is a foundation stone of the 'soft' social sciences and well equipped to guide the construction and presentation of these cases.

It would appear that current evaluation in GTZ goes a long way to meeting the evaluation requirements for INRM. Firstly, the GTZ approach explicitly acknowledges that innovation is a social process. Secondly, the idea of letting projects do their own monitoring and evaluation matches well with the INRM emphasis on adaptive management (Sayer and Campbell, 2001). Researchers cannot have perfect knowledge of either socioeconomic or ecological systems in complex environments,

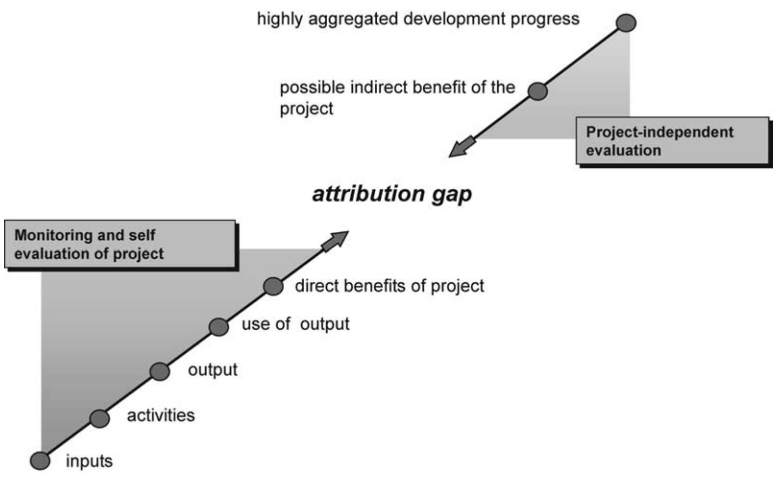

Fig. 3. The GTZ impact model (Kuby, 1999).

hence the sensible approach is management that 'learns by doing' and adapts as it goes along.

The GTZ evaluation model can better meet the requirements of INRM, however, if work to bridge the attribution gap is started during the project life-span and not left completely to the ex post impact assessment. After all, it is easier to build a bridge from two sides that meet in the middle, than solely from one bank.

One advantage of starting to build the bridge at the beginning of the project is that it can help in assessing the likelihood of success of the project and in identifying likely scaling out and up pathways (see Fig. 2). If these pathways are identified early on then they will help guide the project, the stakeholder analyses, and the partnerships that are formed during the project. Secondly, thinking about ex post impact assessment from the outset of a project should lead to base-line surveys being carried out, without which ex post impact assessment is much harder.

3.2. Impact pathway evaluation (IPE)

A program theory is an explicit theory or model of how a project will, or has, brought about impact (Rogers et al., 2000). A program theory consists of a sequenced hierarchy of outcomes—in other words an impact pathway with milestones on the route. The hierarchy begins with the project outputs, followed by a chain of intermediate outcomes that are then followed by the wider and often longer-term outcomes. It represents a set of hypotheses about what needs to happen for the project output to be transformed, over time, into impact on highly aggregated development indicators. It is the hypothetical bridge between project outcomes and eventual impact. Program theory evaluation (PTE) is simply an evaluation guided by the impact pathway.

We call our application of program theory evaluation, guided by the GTZ Impact model, impact pathway evaluation (IPE) because the term impact pathway is better understood in agriculture than program theory. IPE is a two-phase evaluation. In the first phase a project uses PTE to guide its self-monitoring and evaluation. The second phase is an independent ex post impact assessment that would normally be carried out several years after the project has finished. It is this impact assessment that attempts to bridge Kuby's attribution gap, using the phase 1 evaluation results as a foundation.

The particular PTE approach that we think most appropriate for phase 1 is the program theory matrix approach that has been used in Australia for evaluating many different types of programs for over 15 years (Funnel, 2000). However, other PTE approaches may be more suitable depending on needs. In a program theory matrix evaluation a series of questions are posed for each identified outcome in the impact pathway. The answers to these are recorded in a matrix which is similar in many respects to a log frame. Indeed this is an advantage of the program theory matrix approach because most people in the CGIAR system are familiar with log frames. The questions are:

1. What would success look like?
2. What are the factors that influence the achievement of each outcome?

3. Which of these can be influenced by the project?
4. Which factors are outside the direct influence of the project?
5. What is the program currently doing to address these factors in order to bring about this outcome?
6. What performance information should we collect?
7. How can we gather this information?

In the next section we examine one case study in which IPE is being used on purpose, and a second where CIP researchers developed and used an IPE approach without recognising it as such.

3.3. Case Study 1: application of impact pathway evaluation to integrated Striga control (ISC)

Striga hermonthica is a parasitic weed that attaches itself to the roots of cereals (e.g. maize, sorghum, millet and rice), diverting essential nutrients and leaving the host stunted and yielding little or no grain. The weed is the severest biological constraint to cereal production in sub-Saharan Africa, infesting almost 21 million hectares of land causing millions of dollars of damage (Sauerborn, 1991). Farmers world-wide call it 'witch' weed, because it does most of its damage before it emerges from the soil.

Research at IITA and elsewhere is showing that *Striga* control is possible using an integrated approach that attacks *Striga* from several sides. Since 1999, one of the authors has been working in four villages in Northern Nigeria using participatory research approaches to develop locally-adapted integrated *Striga* control (ISC). The villages were chosen on the basis of having severe *Striga* problems. Two group meetings were held, first to carry out a problem consensus to rank *Striga* in relation to other problems, and then to design experiments to evaluate the options for *Striga* control. The R&D team has provided training to improve farmers' understanding of *Striga*. The work began with 19 participating farmers (Schulz et al., 2003).

A key technology in the ISC options that the researchers brought to the villages is the use of a legume crop (e.g., soybean, cowpea, groundnut) that induces a high proportion of *Striga* seeds to germinate, which then die because they cannot parasitize legumes. This is called 'trap cropping', and the discovery that the high genetic diversity of *Striga* requires screening of legumes to find effective trap crops for different localities has been one of IITA's more important research contributions. Another research breakthrough has been the development of *Striga*-resistant maize varieties.

To be effective, legume trap crops must be planted much more closely than farmers usually plant their legumes, and should be planted together with *Striga*-resistant cereals, seed cleaning to remove *Striga* seed, crop rotation, weeding of the *Striga* plants before they set seed, and improved soil fertility.

Fig. 4 shows an impact pathway that we have developed to describe how we expect the project output—on-farm research to adapt and validate ISC options in farmers' fields—might lead ultimately to the project goal of improved livelihoods for the 100 million people in Africa that are affected by *Striga*. The shaded and numbered boxes are the outcomes that we are monitoring. The project purpose, Box 4, is equivalent to the 'direct benefits of the project' shown in the GTZ impact model. The unshaded boxes will be evaluated in a future ex post impact assessment.

After constructing the impact pathway, the next step in program theory matrix approach is to construct a matrix for each of the outcomes shown in these shaded boxes. An example of the matrix for the outcome in Box 4 is shown in Table 1.

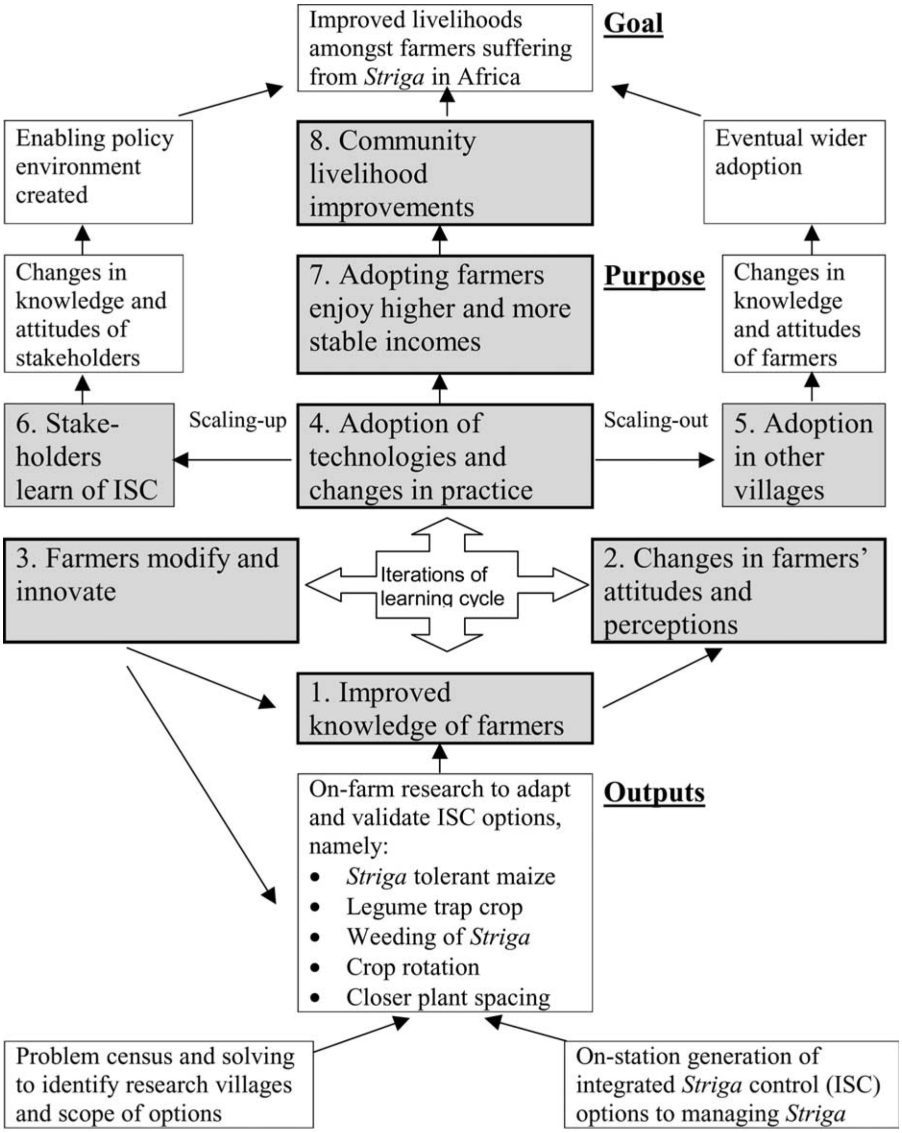

Fig. 4. Impact pathway for an integrated *Striga* control (ISC) Project in Northern Nigeria.

Table 1
The program theory matrix for an intermediate outcome of the Integrated *Striga* Management project

1. Intended outcome	2. Success criteria	3. Program factors affecting success	4. Non-program factors affecting success	5. Activities and resource of program	6. Performance information	7. Sources of data
Farmers modify and innovate to better match the best-bet options to their circumstances	Participating farmers make changes that improve the technology for them, they continue with them and promulgate them to others Non participating farmers do the same Farmer-specific examples from both groups of innovation that has made important improvements On-farm experiments are adapted to reflect this new 'best practice' (i.e., we learn from the farmers).	Extent to which project is working with farmers who have a real problem with Striga Extent to which project explains the on-farm experiments to farmers The visual performance of the ISC package compared to farmer practice Extent to which on-farm experiments are a burden to participating farmers Project efforts made to inform and include farmers who originally choose not to participate Incentives provided to farmers to participate Efforts made to encourage and nurture innovation	Farmers' previous experience with research projects and their subsequent expectations of this one Farmers' beliefs and past experiences with Striga and ISC options Attitudes of the community to innovative behaviour and efforts to better oneself Degree of social inter-connectedness and communication Drought Conflicts and political instability	Use of problem census and problem solving approach to initially agree on research approach and participating farmers Setting up of experimental trials in farmers' fields Farmers' group meetings at the beginning and end of the season to explain, clarify and seek feedback Field technicians regularly visit the research villages to do the same. Group training to explain certain aspects (e.g., Striga life-cycle)	Percentage of farmers who make modifications of any kind Percentage of farmers who make improvements (as judged by M&E team) Examples of major improvements (column 2) Percentage of farmers who maintain their changes Percentage of farmers who promulgate their changes to others Reasons for making modifications, selection and promulgation decisions Feedback on time and effort needed to participate in on-farm experiments	Project documentation including trip reports and report of problem census and problem solving workshops GIS mapping of adoption in farmers' fields followed by individual interviews Group interviews Field technicians' reports Video and photographs of modifications Site visits and case studies by M&E team

Table 1 shows that the PTM approach is similar to the log frame approach in some of the questions asked at each level in the outcomes hierarchy. Importantly, though, the hierarchies are different. The log frame hierarchy is limited to activities-outputs-purpose-goal and a weakness is in the lack of detail as to how the project outputs achieve the project purpose, that is, in terms of the GTZ impact model, how a project's outputs bring direct benefits. It is exactly this area that the PTM maps out in detail. The program theory (impact pathway) shows the intermediate outcomes required for project outputs to bring real benefits (purpose and goal). The matrix helps identify activities and resources that the project can employ to bring achieve the intermediate outcomes, as well as identifying important assumptions.

We are using two published approaches to monitor and evaluate the delivery of the intermediate outcomes shown in impact pathway. The first is the 'follow-the-technology' approach (Douthwaite et al., 2001b; Douthwaite, 2002) that sees technological change in general, and early adoption in particular, as an evolutionary process in which stakeholders generate novelties (i.e., make modifications; innovate), select those that appear to work and spread (i.e., scale up and out) the results.

The follow the technology approach involves, as the name suggests, following new technologies and knowledge as they are adopted. Because it is based on an evolutionary view of the innovation process, the follow-the-technology approach focuses on identifying modifications, selection decisions (i.e., whether farmers decide to adopt a modification), and scaling out and up processes. Key to the direction and nature of an evolutionary process is the environment, hence the follow-the-technology approach pays particular attention to seeking explanations for novelties generated, selection decisions made and the nature of scaling paths to understand the socio-economic and cultural factors affecting stakeholders' learning and decision making processes.

From October 2001 to January 2002 a survey was carried that sought to identify the extent to which participating farmers had increased the use of ISC on their own farms, as well as the extent of farmer-to-farmer diffusion. A total of 245 expansion farmers using one or more ISC methods were identified, and their fields mapped using a hand-held geo-positioning system (GPS). The experimental and expansion plots of the 44 participating farmers were also mapped. A data sheet was completed to record what was planted in the fields, and modifications made to the recommended package shown in Fig. 4. From February to June 2002 an in-depth survey was then carried out of a random sample of 155 of the participating and expansion farmers. The position of the farmers' households were also mapped. The survey established farmers' existing *Striga* control practices and sought explanations for farmers' adoption and modification decisions, his or her understanding of ISC, and to find out where the farmer received the technologies from, and who he or she has passed them on to. The questionnaire specifically asked whether farmers passed on any of the agronomic recommendations, e.g., close legume spacing, in addition to distributing seed. In this way, the FTT approach monitored and evaluated changes to Boxes 1–5 of the outcomes shown in Fig. 4.

Table 2 shows the adoption of the ISC technologies. It shows large differences exist between villages, reflecting differences in farmer preference resulting from differing agro-ecological and socio-economic conditions. This supports the idea that solutions to complex problems do need to be constructed in situ, and that conversely a ‘one size fits all’ solution will not work.

The ISC ‘best bet’ options include the recommendation that farmers grow the legume trap crop as a mono-crop. This is because if cereals are grown with the legumes, then the *Striga* growing on that cereal can replace the seeds killed by the trap crop. Table 3 shows that well over half of farmers rejected the mono-crop recommendation by continuing with their mixed cropping practices. This was initially a concern to the project. However, the survey shows that 81% of farmers had adopted the practice of weeding out *Striga* before it set seed thus reducing the problem of planting cereals with legumes. Most farmers have a strong preference to continue with mixed cropping because it allows them ‘‘to harvest double’’ and to ‘‘guard against crop failure’’. In response, the project is now placing more emphasis on weeding *Striga* and less on mono-cropping legumes than the ‘best-bet’ ISC originally introduced to farmers. The project is also now promoting the modification made by a single farmer of planting two rows of soybean on one ridge. This allows farmers to achieve the necessary close legume spacing required to induce suicidal germination in a high proportion of *Striga* seeds, while at the same time allowing

Table 2
Percentage adoption of ISC technologies by farmers in four pilot villages in Northern Nigeria

Technology	Adoption per village (%)				Average adoption
	Ankwa	Kaya	Mahuta	Rimau	
Soybean trap crop	77	93	71	100	89%
Weeding of Striga	53	82	90	88	82%
Rotation of legume and cereal	41	84	94	83	81%
Striga resistant maize	6	41	55	48	42%
Sole crop of legume	24	11	26	23	19%
Cowpea trap crop	6	2	19	4	7%
Sole crop of cereal	6	4	3	6	4%
Average number of technologies adopted	2.12	3.20	3.58	3.50	3.24
N	17	56	31	48	152

Table 3
Modifications made to researcher-recommended Integrated *Striga* Control (ISC) package

Modification	f	%
1. No modification	58	43
2. Planting widely spaced single rows of cereal in soybean perpendicular to the ridges (Gicci)	39	29
3. Wider row spacing	14	10
4. Strip cropping (e.g. 2 rows cereal, 4 rows legume)	14	10
5. Intercropping (e.g. maize, sorghum and groundnut in same field)	8	6
6. Planting two rows of soybean on 1 ridge	1	1

$N = 134$.

them to continue to use the local animal-drawn plough which has a fixed, but wide, ridge spacing.

Another finding from the M&E was that 93% of farmers were saving at least one of the ISC seed varieties and nearly two thirds of these farmers were giving seed, mainly soybean, on average to 2.3 other farmers. In terms of Fig. 2 the seed components of ISC are scaling-out. Of the farmers who gave seed, 85% also gave advice on how to grow the crop so as to control *Striga*, showing that the crop management knowledge necessary for ISC is also spreading.

Other useful feedback from the M&E is the pattern of adoption shown in Fig. 5, which shows adoption generally clustered around the participating farmers' households. However, in all villages adoption has 'jumped' and new clusters have formed. The project is now carrying out further research to better understand how these jumps occur, and will build an extension approach based on this understanding.

We are monitoring scaling-up (Box 6 in Fig. 4) by keeping a record of the project's interactions with other stakeholders. The follow-the-technology approach does not lend itself to monitoring whether adopting farmers are enjoying higher incomes (Box 7), and whether there are livelihood improvements in the four villages (Box 8). Instead we have constructed case studies of individual households, guided by the sustainable livelihoods framework (SLA) (Scoones, 1998), described in detail at

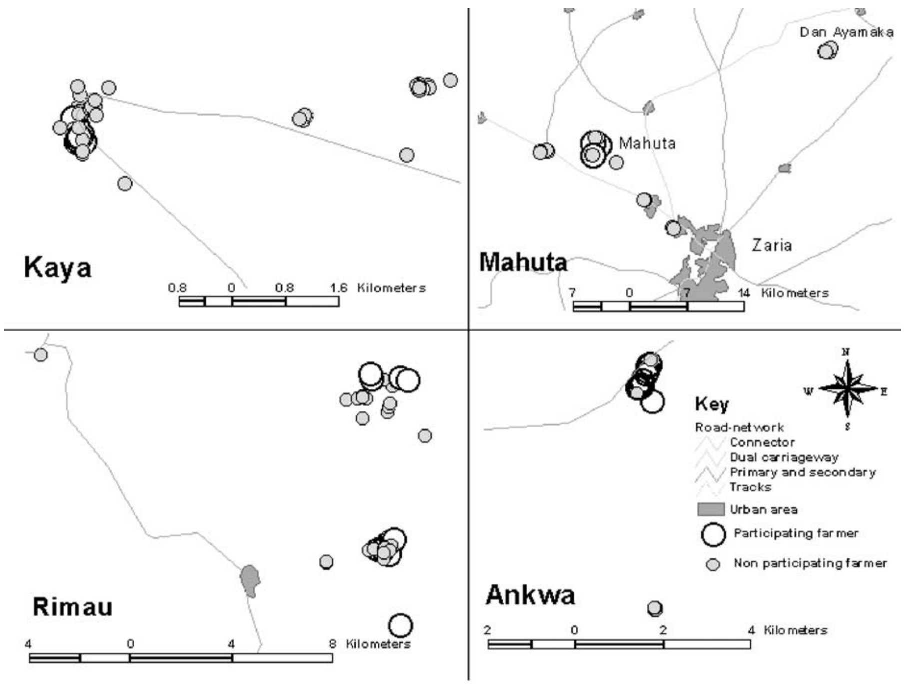

Fig. 5. Adoption and spread of integrated *Striga* control (ISC) technologies in four villages in Northern Nigeria, two (Ankwa) or three seasons after establishment of the first trials.

(http://www.livelihoods.org/index.html). The households were purposively selected to be representative of poor, medium and rich households in the four villages.

3.4. Case study 2: monitoring and evaluation of a participatory sweet potato integrated crop management project in Indonesia

The International Potato Centre (CIP) implemented a project in Indonesia in 1994–1997 to develop an integrated pest management (IPM) approach for sweet potato cultivation. The project, covering a period of six cropping seasons, began with a needs and opportunity assessment that led to the project team concluding that farmers' needs were broader than just pest management. Accordingly, the team widened the project's scope to the development of an Integrated Crop Management (ICM) approach that would be promoted amongst sweet potato farmers using farmer field schools (FFS), originally developed for promoting Integrated Pest Management in rice. The project set up a number of farmer researcher teams (FRTs) in each of four project sites, consisting of two farmers per team.

The FRTs were encouraged to conduct a series of experiments to test and adapt a basket of pest and crop management options. As well as making adaptations the FRTs developed some innovative practices. At the same time researchers carried out additional on-station and on-farm trials, mainly on sweet potato weevil management, because at the time that was the main research interest of CIP's global sweet potato Integrated Pest Management agenda. Farmers and researchers collectively evaluated all FRT and researcher-managed on-farm and on-station trials during 6-monthly end-of-season evaluation and planning workshops. The ICM guidelines were developed and, during later seasons, adapted on the basis of these evaluations, and knowledge gaps identified set the next season's research agenda.

Simultaneously, the team took steps to develop a protocol for large-scale promotion of the ICM approach. A sweet potato ICM FFS curriculum was drafted and experiential learning modules developed. The ICM modules were pre-tested during two seasons on groups of sweet potato farmers and the FRTs. The project team monitored and evaluated these FFS to identify ways of improving the ICM guidelines and the FFS modules and curriculum. This M&E used success indicators defined by the participating farmers (van de Fliert et al. 2001). At the end of the project (late 1997), the sweet potato ICM guidelines and FFS modules were published in Indonesian in a manual (Van de Fliert and Braun, 1999).

In order to prepare for large-scale sweet potato ICM FFS implementation programs by extension organisations, the project involved two batches of officials and farmer trainers in two-week training-of-trainers events. One batch of 40 trainees came from the government National Integrated Pest Management Program and the other 42 trainees from local NGOs. The final task in the training-of-trainers consisted of the development of a follow-up work plan for sweet potato ICM FFS implementation, including funding plans. The National Integrated Pest Management Program decided to allocate their special "Follow-up FFS" funds (for rice Integrated Pest Management Program FFS alumni) for one round of sweet potato ICM FFS in the six major sweet potato growing districts in Java, and the Directorate for Root crops contributed some additional funds for transportation fees for the farmer trainers to be heavily involved. These FFSs were implemented during the 1997-1998 wet and dry season, and were anticipated to trigger interest from local (district and province) governments to scale the process up during consecutive seasons. However, Indonesia was hit by a serious economic crisis in 1997 and many government development programmes collapsed, including continuation of sweet potato ICM FFS. Nevertheless, some of the trained farmer groups continued on a self-supported basis with collective experimentation and expansion of sweet potato FFS within their communities.

The NGOs had a more diversified set of follow-up activities to the training-of-trainers in order to accommodate their ongoing community development programmes with their constituencies. Some did organise sweet potato ICM FFSs, but others conducted ICM FFS in other crops, or incorporated the ICM principles and experiential learning approaches in their existing activities, which has continued to date.

The CIP project team received funding from UPWARD[1] to evaluate the impact of the sweet potato ICM FFSs conducted by the National Integrated Pest Management Program during 1998–1999. UPWARD also supported the evaluation of the NGO programmes during 2001–2002 with a small grant to one of the NGO networks trained previously. The approach applied in both evaluations used a framework developed by a team of stakeholder representatives and built on the M&E indicators defined during the FFS development stage (van de Fliert et al., 2001). In 2000 an external case study review was carried out by the CGIAR system's Special Program on Participatory Research and Gender Analysis (SP-PRGA) which focused on the contribution of farmer participation in the impact achieved by the FFS (Johnson et al., 2001).

Although the project did not know of PTE, the basis of both the internal and external evaluations was a conceptual model showing the project outputs leading to the direct benefits (impact on farmers' livelihoods in the pilot areas), i.e., an impact pathway showing a sequenced hierarchy of outcomes. This model is shown in Fig. 7. And again, despite the project not knowing of program theory matrix evaluation, the first step in the M&E exercises was to identify what success would look like, based on the principles and objectives written into the ICM and FFS protocols, and then derive evaluation variables and indicators that would reflect the level of success. In other words the M&E approach explicitly covered columns 1, 2 and 6 of the program theory matrix shown in Table 1. Indeed, further similarity with the integrated *Striga* control case study is that the project considered aspects of the sustainable livelihoods framework by ensuring that the evaluation indicators covered changes in four of the five capitals—human, social, environmental (natural) and financial. The project also followed the GTZ-recommended practice of carrying out its own self-evaluation and this clearly contributed to the success of the work.

[1] User's Perspective with Agricultural Research and Development, which is a CIP-affiliated network of Asian researchers working on participatory approaches in root crop systems. UPWARD also partly funded the first sweet potato ICM FFS development project in 1994–1997.

The second step was to identify sources of information for the various project phases shown in Fig. 6, thus covering column 7 of the program theory matrix shown in Table 1. These information sources are shown in Table 4.

In addition to covering columns 1, 2, 6 and 7, the M&E of the ICM project also explicitly considered the activities and resources of the program dedicated to achieving the different outcomes. Columns 3 and 4—consideration of program and

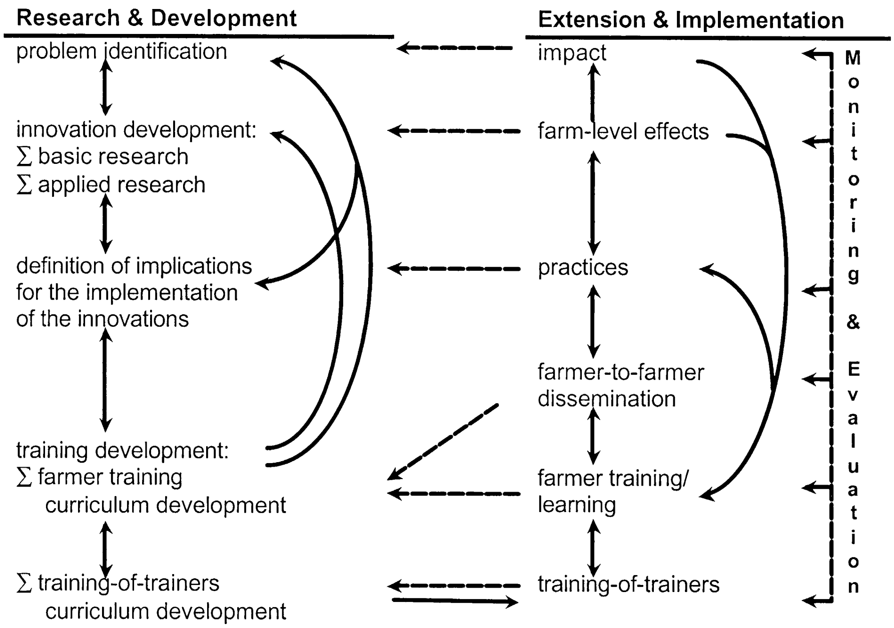

Fig. 6. Framework used for planning, monitoring and evaluation of an Integrated Crop Management (ICM) project in Indonesia (Van de Fliert and Braun, 2002).

Table 4
Information sources for the various evaluation phases of the sweet potato ICM project (after van de Fliert et al., 2001)

Information source	Evaluation phase					
	Baseline (pre-FFS)	Training-of-trainers	FFS implementation	Farmer-to-farmer dissemination	Post-FFS ICM implementation in farmers' fields	Post-FFS effects at farm level
Trainee/ ICM farmer	✓	✓	✓	✓	✓	✓
Non-ICM farmer	✓			✓	✓	✓
Trainer/ facilitator		✓	✓	✓	✓	
Sweet potato field	✓		✓		✓	✓
Village official	✓		✓			
Trader	✓					✓

non-program factors affecting the success of the project—were taken into account through informal interviews and discussions during the final analysis workshop. If the project had known about PTM evaluation prior to the evaluation then the M&E would have considered factors affecting the success of the project, and how those might have been influenced by the project, more systematically and earlier. This would have helped with project implementation.

What is innovative about CIP's M&E approach, and of particular interest to INRM, given the previous discussion about the importance of scaling, is the division of the program theory into R&D component and a scaling component, and the recognition that M&E of the scaling part is important in validating and improving the R&D. This is in keeping with Lyotard's (1984) idea that the validation of knowledge is whether it is useful to the intended beneficiaries. This feedback is depicted as the horizontal dashed lines in Fig. 6. This makes clear that a project such as this is not a linear progression from research to extension but an iterative learning process that adapts as it goes along. This is the key to good adaptive management which is a core principle of INRM.

In practice this impact-pathway-like evaluation approach proved very useful. In addition to helping guide M&E, as already discussed, the evaluation found that farmers can grow sweet potato more sustainably by applying ICM, and that the FFS provides an appropriate platform for farmers to effectively learn, experiment and organise themselves (van de Fliert et al., 2001). Additionally, the PRGA study showed that in this case farmer participation in research did make a difference to achieving impact (Johnson et al., 2001).

3.5. Discussion of the case study findings

Both the integrated *Striga* control (ISC) and the integrated crop management (ICM) impact pathways concentrate largely on the left-hand-side of the GTZ impact model, that is, in detailing the outcomes that link the project outputs to the direct benefits to the target groups in the pilot areas where the projects have worked. The intended direct benefits of both the ISC and ICM projects include improving farmer profitability and incomes. However, both projects have gone, or will go, further by assessing how these tangible benefits are impacting on the livelihoods of farmers in the communities in which the projects work.

Beginning to span the attribution gap in this way during the self-evaluation stage will make the second stage ex post impact assessment much easier. For example, the ICM project can claim wider impact if it can show that: (1) FFSs have been carried out in other areas and livelihood improvements are measured in these areas; (2) these improvements happened after the FFSs; (3) there is a plausible explanation of how FFSs led to these livelihood improvements; and (4) if alternative explanations can be discounted. The project M&E has helped establish three of these proofs because it has shown in the pilot areas what types of livelihood impacts are likely, has given an explanation of how the project outputs led to these impacts, and has considered, and discounted, alternative explanations for the improvements (Johnson et al., 2001).

A second advantage of considering how a project is going to achieve wider impact from the outset is that steps can then be taken to help bring it about. For example, in the ISC project we may well find that wider impact will require scaling-up in the form of the extension services in Nigeria adopting a new approach to *Striga* control extension. This will require a directive from the Federal Ministry of Agriculture. Hence, the project might decide to be proactive and commission an ex ante assessment of the benefits of *Striga* control, aimed at key individuals in the Ministry. In terms of Fig. 2 this would represent the project making an intervention at the national government level.

In general, though, both the ISC and ICM projects place greater emphasis on fostering scaling-out, that is, grassroots spread of at least some of the project's outputs, because this is a necessary condition for scaling-up. The ICM case study showed that plans for scaling-up are very prone to external shocks, that is, factors outside any project's control.

Fostering scaling out and up (Fig. 2) is best done by first identifying who are the key stakeholders—the people who will ultimately benefit from the innovations and the people responsible for their promulgation—and then working with these stakeholders in a participatory way to encourage them to take over ownership. If this happens then the key stakeholders will tend to promote it to each other and lobby for political support for the work, even if there are setbacks and funding cuts.

4. Conclusions

Research aimed at sustainably improving rural peoples' livelihoods is largely based on the constructivist paradigm, which is different to the positivist one that was used as the basis of the research that successfully catalyzed the Green Revolution. Hence, such research is conducted with a different set of assumptions, concepts, values, and practices, that also require a different type of evaluation approach, than the economic ones that represent current 'best practice' in the CGIAR system. Constructivist-based research acknowledges that technological change is a highly social process that makes it practically impossible to produce a systematic causal relationship between the costs and general benefits of research that conventional economic assessment methods require. This paper finds that impact pathway evaluation (IPE), based on the GTZ impact model and program theory evaluation (PTE), is better matched to the needs of research that considers people. Where conventional economic assessment methods are possible, IPE can be highly complementary by providing process understanding and a human dimension to the quantitative analysis.

In the first stage of IPE a research project begins by developing an impact pathway for itself and then conducts a self-evaluation, guided by the impact pathway, to the point of establishing the direct benefits of its outputs in its pilot site(s). Self-evaluation, and the learning it engenders, is necessary for successful project management in complex environments. The impact pathway also evolves, based on this learning, to map out in greater detail how the project's direct benefits are likely to lead to wider impact.

The second stage is an ex post impact assessment some time after the project has finished, in which the project's wider benefits are independently assessed, paid for by a coalition of donors with interest in the area in which the work had been done. This assessment begins by establishing the extent to which the impact pathway was valid in the pilot sites and the extent to which scaling has occurred. It is the job of the impact assessor to build a plausible and persuasive case for a link between the project outputs and general level developmental changes using case study methodology together with quantitative approaches, if possible.

The program theory matrix approach, which can be part of an IPE, describes how the project outputs are expected to produce direct benefits to the stakeholders involved. While similar in some respects to a log frame, the program theory matrix approach concentrates on an area of log frame weakness, which is the jump from project outputs to direct benefits (project purpose). The approach does not describe the jump from direct benefits to goal (i.e., contribution to achieving highly aggregated development objectives) in such detail, but it does identify steps the project should take to scale out and up. Taken together, this emphasis on detailing the activities and resources needed to turn project outputs into direct benefits to the intended beneficiaries, and then how wider impact might then be achieved, makes eventual impact more likely. It is an aid to experiential learning for all stakeholders concerned.

Acknowledgements

An earlier version of this paper was presented at the international conference "Why Has Impacts Assessment Research Not Made More of a Difference?", held 4-7 February, 2002 in San José, Costa Rica. The conference was co-organized by the Standing Panel on Impact Assessment (SPIA) of the Consultative Group for International Agricultural Research (CGIAR) and by the Economics Program of the International Maize and Wheat Improvement Centre (CIMMYT).

References

Alston, J.M., Norton, G.W., Pardey, P.G., 1995. Science under Scarcity. Cornell University Press, New York, USA.

Anon, 2002. Current definition of INRM. 1st March, downloaded from (http://www.inrm.cgiar.org).

Bijker, W., Pinch, T., 1988. The social construction of facts and artifacts: or how the sociology of a science and the sociology of technology might benefit each other. In: Bijker, W., Hughes, T., Pinch, T. (Eds.), The Social Construction of Technology Systems. MIT Press, Cambridge, Mass., USA.

Collinson, M.P., 2000. Introduction. In: Collinson, M.P. (Ed.), A History of Farming Systems Research. FAO, Rome and CABI Publishing, Wallingford, England, pp. 1–4.

Douthwaite, B., 2002. Enabling Innovation: a Practical Guide to Understanding and Fostering Technological Change. Zed Books, London, England.

Douthwaite, B., de Haan, N.C., Manyong, V., Keatinge, D. 2001a. Blending "hard" and "soft" science: the "follow-the-technology" approach to catalyzing and evaluating technology change. Conservation Ecology 5(2): 13. [online] URL: (http://www.consecol.org/vol5/iss2/art13).

Douthwaite, B.J., Keatinge, D.H., Park, J., 2001b. Why promising technologies fail: the neglected role of user innovation during adoption. Research Policy 30 (5), 819–836.

Funnel, P. 2000. Developing and using a program theory matrix for program evaluation and performance monitoring. In: (Rogers, P.J., Petrsino, A., Hasci, T.A., Huebner, T. (Eds.), Program Theory in Evaluation. Challenges and Opportunities. New Directions for Evaluation 87.

GAO (General Accounts Office), 1987. Case study evaluation. Program Evaluation and Methodology Division, Transfer Paper 9. GAO, Washington DC, USA.

Horton, D., 1998. Disciplinary roots and branches of evaluation: some lessons from agricultural research. Knowledge and Policy 10, 4.

Houghton Mifflin Company 2000. The American Heritage Dictionary of the English Language (on-line). fouth ed. Downloaded from (http://www.eref-trade.hmco.com/).

IIRR (International Institute of Rural Reconstruction). 2000. Going to scale. Can we bring more benefits to more people more quickly? Highlights of a workshop held at the IIRR, Silang, Cavite, Philippines. Downloaded from (http://www.ngoc-cgiar.clades.org/scalin1.htm).

Johnson, Nancy, van de Fliert, Elske, Lilja, Nina, 2001. The CIP development of integrated crop management (ICM) technologies and practices for farmer field school (FFS) for sweet potato in Indonesia (1990s). In: Johnson, N., Lilja, N., Ashby, J. (Eds.), Characterizing and Measuring the Effects of Incorporating Stakeholder Participation in Natural Resource Management Research: analysis of Research Benefits and Costs in Three Case Studies. PRGA Working Document No. 17. Cali, Colombia pp. 21–54.

Kuby, T., 1999. Innovation as a Social Process: What Does this Mean for Impact Assessment in Agricultural Research. Paper presented at a CIAT workshop, Costa Rica, September 1999.

Kuby, T. 2000. Turning Attention Towards Results: how GTZ is Building its Impact Evaluation Capacity. Internal GTZ document. Eschborn, Germany.

Kuhn, T.S., 1970. The Structure of Scientific Revolutions, revised ed.. The University of Chicago Press, Chicago, USA.

Leeuw, F. 2000. Program Evaluation and Social and Institutional Impact Assessment. Paper presented at "The Future of Impact Assessment in the CGIAR: needs, Constraints and Opportunities" Workshop. Standing Panel on Impact Assessment (SPIA) and the Technical Advisory Committee (TAC) of the CGIAR, FAO, Rome, 3–5 May.

Lyotard, J.F., 1984. The Post-modern Condition. (Trans. G. Bennington and B). Massumi. University of Minnesota Press, Minneapolis.

McCown, R.C. 2001. Farming systems research and practice. Proceedings of the 10th Australian Agronomy Conference, Hobart, Australia.

Rogers, P.J., Petrsino, A., Huebner, T., Hasci, T.A. 2000. Program theory evaluation: practice, promise, and problems. In: (Rogers, P.J., Petrsino, A., Hasci, T.A., Huebner, T.) Program Theory in Evaluation. Challenges and Opportunities. Eds. New Directions for Evaluation 87.

Rogers, E.M., 1995. Diffusion of Innovations, fourth ed. (Rev. ed. of: communication of Innovations). The Free Press, New York.

Röling, N., 1988. Extension Science: Information Systems In Agricultural Development. Cambridge University Press, Cambridge.

Röling, N., 1995. Towards an interactive agricultural science. Journal of Agricultural Education and Extension 4, 35–48. (http://www.agralin.nl/ejae/v2n4-5.html).

Rosenberg, N., 1982. Inside the Black Box: Technology and Economics. Cambridge University Press, Cambridge.

Sauerborn, J., 1991. The economic importance of phytoparasites *Orobranche* and *Striga*. In: Ransom, J.K., Musselman, L.J., Wosham, A.D., Parker, C. (Eds.), Proceedings of the 5th International Symposium on Parasitic Weeds, 24–30 June 1991. International Maize and Wheat Improvement Centre (CIMMYT), Nairobi, Kenya, pp. 137–143.

Sayer, J., Campbell, B. 2001. Integrated natural resource management research to integrate productivity enhancement, environmental protection and human development. Conservation Ecology 5(2) (http://www.consecol.org/vol5/iss2/).

Schulz, S., Hussaini, M.A., Kling, J., Berner, D.K., Ikie, F.O., 2003. Evaluation of integrated *Striga hermonthica* control technologies under farmer management. Experimental Agriculture 39, 1–10.

Scoones, I., 1998. Sustainable Rural Livelihoods: A Framework for Analysis. IDS Working Paper 72. University of Sussex, Brighton.

Van de Fliert, E., Braun, A.R. 1999. Farmer Field School for Integrated Crop Management of Sweet potato: Field Guides and Technical Manual. International Potato Centre, Bognor, Indonesia.

Van de Fliert, E., Braun, A.R., 2002. Conceptualizing integrative, farmer participatory research for sustainable agriculture: from opportunities to impact. Agriculture and Human Values 19 (1), 25–38.

Van de Fliert, E., Johnson, N., Asmunati, R., Wiyanto, X., 2001. Beyond higher yields: the impact of sweet potato integrated crop management and farmer field schools in Indonesia. CIP Program Report 1999–2000, 331–342.

Yin, R.K., 1994. Case Study Research: Design and Methods, second ed. Sage, London.