Discussion Paper

No. 96-7 **April, 1997**

Disciplinary Roots and Branches of Evaluation: Some Lessons from Agricultural Research*

Douglas Horton

International Service for National Agricultural Research (ISNAR)
P.O. Box 93375
2509 AJ The Hague
The Netherlands
Tel: (31) 70-3496100
Fax: (31) 70-3819677
E-mail: isnar@cgiar.org
Internet: www.cgiar.org/isnar

* Edited version. Published in Knowledge, Technology, and Policy. Volume 10, November 4. Winter 1998. Pages 31-188.

Disciplinary Roots and Branches of Evaluation: Lessons from Agricultural Research

Douglas Horton[**]

Abstract

Since its origins in North America in the 1960s, the field of program evaluation has grown considerably and its concerns have broadened from accountability to program improvement, decision support and institutional learning. Program evaluation is now commonly practiced in governmental organizations not only in North America but also in many countries of Western Europe and Oceania. Although program evaluation is a relatively new field with many controversies and lively debates, a unifying body of evaluation theory, methods and standards is gradually emerging. Evaluation has recently been described as a "transdiscipline" like statistics and measurement.

This paper is based largely on my personal experiences working in agricultural research organizations in developing regions. Here, a number of different types of evaluation are carried out, but program evaluation (as defined by Patton, 1997) is largely unknown. Distinct branches of agricultural research evaluation can be identified, with disciplinary roots in the natural sciences and in agricultural economics. The most rigorous agricultural research evaluations are economic studies. Systematic internal evaluation is notably lacking. Current pressures to improve performance, transparency and accountability are creating demands for more systematic evaluation, and many program evaluation concepts and methods would seem to be of value in agricultural research organizations. However, in the current scenario of declining funding for agricultural research managers are hesitant to expand their evaluation activities and explore unfamiliar paradigms and methods. Moreover, they not yet convinced that social-science-based program evaluation would produce useful results. Natural scientists and economists tend to view program evaluation as "soft science" or no science at all.

1. Introduction

Program evaluation is defined in a leading text as:

> "the systematic collection of information about the activities, characteristics, and outcomes of programs to make judgments about the program, improve program effectiveness, and/or inform decisions about future programming" (Patton, 1997).

Modern program evaluation emerged in the U. S. A. in the 1960s to meet demands for accountability and feedback associated with social programs initiated under President Kennedy and expanded under Presidents Johnson and Nixon (Shadish, Cook and Leviton, 1991). Program evaluation has grown substantially over the last 30 years as a field of professional activity and it has spread beyond North America to Oceania, Western Europe and other regions. Professional evaluation associations have been established in several countries including Canada, the United Kingdom, the United States of America and Sweden;

[**]I would like to thank Jock Anderson, Ted Brush, Fred Carden, Willem Janssen, Paul Perrault Warren Peterson and William Shadish for insightful comments on an earlier version of this paper. Ron Mackay offered especially useful suggestions for improving the paper's concluding section. Claudia Forero was responsible for text processing and Jacobine Verhage and Fionnuala Hawes prepared the figures.

Australasian, Central American and European associations have also been founded. Controversies still rage in the associations over issues related to evaluation purposes (e. g. valuing products vs. improving programs), paradigms (scientific vs. constructivist) and methods (quantitative vs. qualitative). Nevertheless, there is a growing core body of generally accepted program evaluation theory and methods. Several specialized evaluation journals are being issued and publishers like Sage Publications market numerous texts, reference books, "how-to" manuals and computer software for evaluation. standards for assessing evaluations have been published. Scores of university-level evaluation courses are available and several universities offer advanced degrees in evaluation.

Michael Scriven, one of the founders of the American Evaluation Association and one of the field's most prolific and authoritative authors, views evaluation as an emerging "transdiscipline" which, like statistics and measurement, can be applied to a range of different primary disciplines and subject matters (Scriven, 1993).

Since 1975 I have worked in international agricultural research centers affiliated with the Consultative Group on International Agricultural Research. The 16 centers affiliated with the CGIAR carry out research and research-related activities on food crops, livestock, forestry, fisheries and natural resource management in the tropics and sub-tropics. Three of the centers were established in the 1960s before the CGIAR was created; most of the centers were established in the 1970s; in the 1980s, the groundwork was laid for a second wave of expansion which occurred after 1990 (Table 1 & Figure 1).

The CGIAR began in the early 1970s as an informal group of donors committed to improving agricultural production in developing countries through research and related activities. As the CGIAR grew and increased in complexity, a Secretariat was established in the World Bank and a Technical Advisory Committee (TAC) set up offices in the Food and Agricultural Organization of the United Nations.

The CGIAR has been characterized as one of the most "over-reviewed" and "over-evaluated" entities in the world. Center programs and management are periodically reviewed by external panels of experts contracted by TAC, and each center carries out an annual internal review of its programs. Individual donors frequently evaluate the projects they support at centers. The CGIAR as a whole was reviewed in 1976. Since then, a number of "stripe reviews" and "stripe studies" have been carried out to evaluate particular fields of research or related work in which a number of centers are involved (for example, root crop research, farming systems research, policy and management research or training carried out by the CGIAR). A second system-wide review is now being organized. In 1984-85 a large-scale exercise was organized to assess the impact of the CGIAR system (Anderson, Herdt & Scobie, 1988). In 1995 the CGIAR established a new, independent Impact Assessment and Evaluation Group (IAEG).

From 1975 to 1990 I was head of the Social Science department of the International Potato Center (CIP). Since early 1990 I have worked as an evaluation specialist at the International Service for National Agricultural Research (ISNAR). Much of my work at CIP and ISNAR has been concerned with evaluation. At CIP, from the earliest days I was involved in participatory, on-farm needs assessment and evaluation of new technologies (Horton, 1984). While I was at CIP, the center was reviewed three times by external panels and many of the social science projects were evaluated by external groups. On two occasions, I led teams that assessed the impact of CIP's programs. Since I moved to ISNAR in 1990, my work – research, training and advisory services -- has focused nearly exclusively on the evaluation of agricultural research activities and organizations.

Table 1. International agricultural research centers affiliated with the CGIAR

Center	Date of foundation (and joining)	Headquarters location	Main commodities and research areas
Original members founded before the CGIAR:			
IRRI (International Rice Research Institute)	1960 (1971)	Los Baños, Philippines	Rice
CIMMYT (Centro Internacional de Mejoramiento de Maíz y Trigo)	1966 (1971)	Mexico City Mexico	Wheat, maize
IITA (International Institute of Tropical Agriculture)	1967 (1971)	Ibadan, Nigeria	Production systems for the humid lowland tropics; rice, maize, cassava, cowpea, plantain, yams, sweet potato
CIAT (Centro Internacional de Agricultura Tropical)	1967 (1971)	Cali, Colombia	Production systems for the humid lowland tropics, rice, beans, cassava, tropical pastures
Founded by the CGIAR or joined the system in the first wave of expansion:			
ICRISAT (International Crops Research Institute for the Semi-Arid Tropics)	1972 (1972)	Hyderabad, India	Production systems for semi-arid tropics; sorghum, millet, chickpea, pigeonpea, groundnut
CIP (Centro Internacional de la Papa)	1970 (1973)	Lima, Peru	Potato, sweet potato
ILRAD [b] (International Laboratory for Research on Animal Diseases)	1973 (1973)	Nairobi, Kenya	Animal disease; theileriosis, trypanosomiasis
ILCA [b] (International Livestock Center for Africa	1974 (1974)	Addis Ababa Ethiopia	Animal feed and production systems; cattle, sheep, goats
IPGRI [d] (International Plant Genetics Resources Institute)	1974 (1974)	Rome, Italy	Plant genetic resources (collection and resources conservation)
WARDA [a] (West Africa Rice Development Association)	1970 (1970)	Bouaké, Côte d'Ivoire	Rice
ICARDA(International Center for Agricultural Research in the Dry Areas)	1975 (1975)	Aleppo, Syria	Wheat, barley, chickpea, lentil, faba beans, pasture and forage legumes
ISNAR(International Service for National Agricultural Research)	1980 (1980)	The Hague, The Netherlands	Research policy organization and management

Table 1.*(Cont.)* **The network of international agricultural research centers coordinated by the CGIAR.**

Center	Date of foundation (and joining)	Headquarters location	Main commodities and areas of focus
Joined the system in the second wave of expansion, after 1990:			
ICRAF [a] (International Centre for Research in Agroforestry)	1977 (1991)	Nairobi, Kenya	Agroforestry; multipurpose trees
IIMI [a] (International Irrigation Management Institute)	1984 (1992)	Colombo, Sri Lanka	Irrigation management
ICLARM [a] (International Center for Living Aquatic Research Management)	1977 (1992)	Manila, Philippines	Fisheries management
INIBAP [c] (International Network for the Improvement of Banana and Plantain)	1984 (1992)	Montpellier, France	Plantain and banana
CIFOR(Center for International Forestry Research)	1993 (1993)	Bogor, Indonesia	Forestry management

a. Joined the Group (i.e. not founded by the CGIAR).
b. Combined to form the International Livestock Research Institute in 1995.
c. Jointed the Group in 1992 and amalgamated with IPRGRI in 1994.
d. Before 1994 known as the International Board for Plant Genetic Resources (IBPGR).

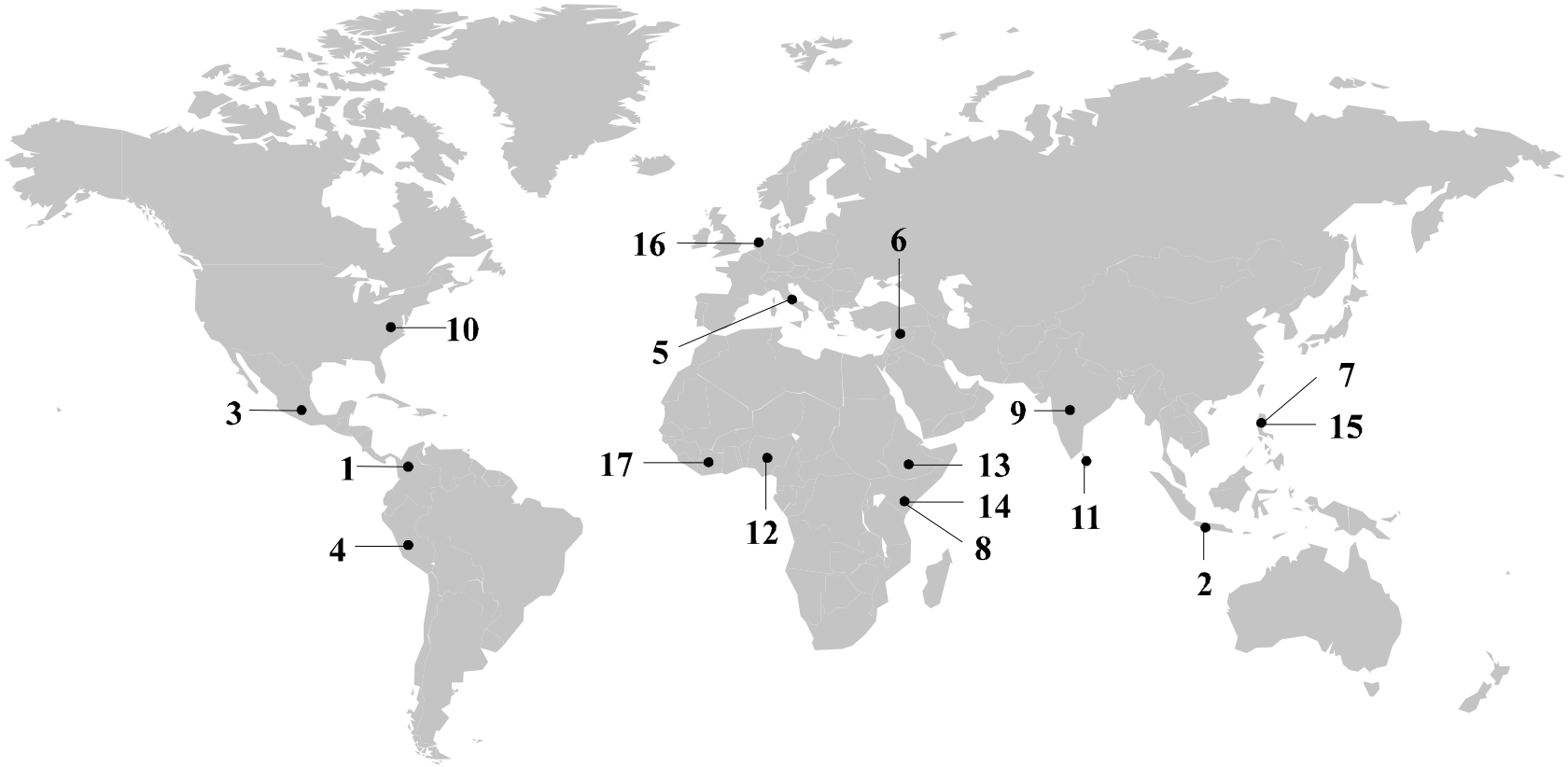
Source: Based onTribe (1994), Table 7.5.

Despite my active involvement in evaluation over the years, I had never heard of "program evaluation" or knew there were such things as professional evaluation associations or specialized evaluation journals until a chance meeting with an evaluator at the World Bank in 1990!

After I "discovered" program evaluation, I began to explore its origins, evolution and the theories and methods developed by evaluation practitioners. Many times I wondered why so little was known in the agricultural research community about program evaluation; it seemed as if these two fields simply never came into contact with one another.

This paper is an attempt to explain why program evaluation has not been applied in agricultural research organizations in developing regions and to assess the prospects for its application in the future. Hopefully, in addition to illuminating the barriers to communication and learning that separate professions, disciplines and sectors, the paper will stimulate experimentation with program evaluation and its successful application in agricultural research organizations.

Figure 1. The CGIR global network. Location of CGIAR-supported International Agricultural Research Centers (IARCs)



1.	CIAT	7.	ICLARM	13.	ILRI
2.	CIFOR	8.	ICRAF	14.	ILRI
3.	CYMMIT	9.	ICRISAT	15.	IRRI
4.	CIP	10.	IFPRI	16.	ISNAR
5.	IPGRI	11.	IIMI	17.	WARDA
6.	ICARDA	12.	IITA		

I begin by identifying some important technical and institutional features of agricultural research that influence evaluation practice. I then describe the main objects of evaluation in agricultural research organizations and the principal methods used to evaluate them. It is shown that the main evaluation methods fall into distinct *branches* which have strong *disciplinary roots* in the natural sciences and agricultural economics. In the final section, I assess prospects for the future and identify some areas in which program evaluators and other key stakeholders could facilitate the application of program evaluation concepts and methods in agricultural research organizations.

2. The context of agricultural research evaluation

Agricultural research in developing countries

Historically, crop and livestock production increased slowly as new lands were brought into production and farmers selected crop varieties and livestock breeds adapted to local conditions. Discovery of the New World and the era of colonization stimulated agricultural growth, opening up new markets and permitting the spread of plants and animals from their historical centers of origin to other parts of the world.

In the nineteenth century, an important institutional innovation took place: Experiment stations and research centers were established and began to apply scientific methods in the development of new agricultural technology.

From its origins in Europe, Japan and North America, scientific agricultural research spread to other parts of the world in the present century (Ruttan, 1982). At present, most countries have some sort of agricultural research system that supports local agriculture through the solution of local production problems and the provision of new crop varieties, livestock breeds and associated production technologies.

Since the Second World War significant investments have been made in agricultural research and extension in Africa, Asia, Latin America and the Caribbean. (Pardey, Roseboom & Anderson, 1991). Beginning in the 1970s, a system of international agricultural research centers known as the Consultative Group on International Agricultural Research (CGIAR) was established (Brady, 1994). The mission of the CGIAR is to contribute to food security, poverty alleviation and conservation of natural resources through applied international research and related development activities (CGIAR, 1995a). Established in 1971 with 3 centers, the CGIAR has expanded to include 16 centers with an annual budget of approximately US $350 million.

The CGIAR has played a significant role in agricultural research, but it is dwarfed by the number, size and scope of national agricultural research organizations in developing countries (Pardey, Roseboom & Anderson, 1991). Brazil's national agricultural research enterprise alone has an annual budget of approximately US $300 million. Each year, the developing countries combined spend at least 20 times as much as the CGIAR system on agricultural research. Furthermore, national agricultural research organizations (NAROs) work on a much wider portfolio of crops and livestock than CGIAR centers (which focus mainly on major food crops, livestock, forestry and fisheries).

Early successes in breeding high-yielding varieties of rice and wheat led to the so-called "Green Revolution" in the 1960s and 1970s (Anderson, Herdt & Scobie, 1988; Tribe, 1994). Since then, progress in crop and livestock improvement has been less spectacular. In recent years, agricultural research organizations have been pressed to move beyond commodity

research and pay greater attention to conserving natural resources and protecting the environment.

Since 1990, the role of public-sector organizations is being reassed and the value of overseas development assistance is being questioned. Public funding for development work is declining in most industrial countries, and funding for agricultural research is being cut disproportionately. As they implement structural adjustment programs and reform the public sector, many developing countries are reducing their public expenditures on agricultural research. These trends have led in many instances to deep cuts in budget and staffing and to profound reassessments of the mandates and priorities of agricultural research organizations.

In light of the shifting public concerns, reduced funding and generally uncertain future, agricultural researchers are eager to improve the management and accountability of their organizations. In this context, they view evaluation as a potentially valuable management tool.

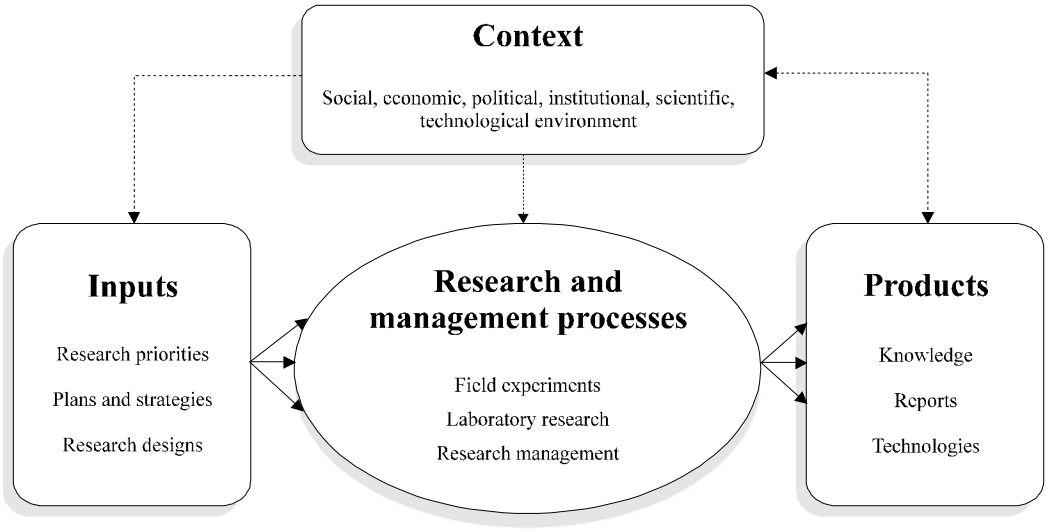
Technical and institutional aspects of agricultural research

A number of technical and institutional features of agricultural research influence the types of evaluation that are undertaken and the methods used. Many models of research organizations and processes have been presented. The one presented here is based on the "CIPP evaluation model," developed by Stufflebeam and colleagues (Stufflebeam, 1984). This model has four main elements, which correspond to the basic components of a comprehensive program evaluation:

> ***Context.*** A program's context includes socio-economic, political, institutional and technological variables in its environment, which influence needs and opportunities . A program's objectives should be evaluated in relation to its context. Were the objectives relevant when the program began? Are they still relevant?
>
> ***Inputs.*** These are the plans and designs on which the program's operations are based. How well do implementation plans correspond to program objectives? Are they adequate for guiding the program and producing the desired outputs?
>
> ***Processes.*** What has the program attempted to do? Have plans been implemented? Has the management of staff, facilities and finances been in accordance with plans and acceptable standards?
>
> ***Products***. What has been achieved? Have planning targets been met? Have expectations been fulfilled? Have program outputs led to significant benefits? Who has benefited and by how much?

Following this model, agricultural research can be viewed as a specialized production *process* involving experiments and related activities that take place in a specific socio-economic *context*, in our case in an agricultural research organizations in a developing country. The research process is guided by *inputs,* in the form of research plans and designs. The process generates *research products*, primarily knowledge and information which is contained in research reports or embodied in new technologies (Figure 2).

Figure 2. The agricultural research process in relation to its context, inputs and products (after Stufflebeam, 1984)



Context and objectives of agricultural research

Most farmers in the developing regions of Africa, Asia and Latin America are smallholders. Many of them have scarce resources and live in or near poverty. Urban poverty is also common and food security is a major issue for many households. Pressure on the land is causing land degradation and threatening environmental sustainability.

Much of the technology needed by small farmers to expand food production and protect the environment takes the form of "public goods," like open-pollinated crop varieties and integrated pest control measures, that cannot be marketed at a profit by private firms. Once they are released, they are freely available to all. Since private firms have little incentive to produce public goods, public support is vital for agricultural research in developing countries.

Historically, most agricultural research has been carried out by public-sector organizations. Private industries engage in R&D to produce new fertilizers, hybrid seeds, pesticides, machinery and other products that can be marketed at a profit. Some large plantations and producers' associations also engage in applied research to improve production and post-harvest methods for export crops like coffee, tea, rubber and sugar cane. But, small farmers cannot afford to support formal research efforts unless there are extraordinary benefits, for example from exports.

For these reasons, this paper focuses on the evaluation of publically funded agricultural research.

Agricultural research funded by the public treasury is carried out to produce information and technologies that contribute to broad socio-economic goals like agricultural growth, food security, poverty reduction and protection of the natural environment. The specific objectives that an agricultural research organization pursues and the relative weight attached to each depends on the specific social, economic and political context. For example, in the 1960s, the mission of the CGIAR focused on productivity; in the 1970s CGIAR centers became more

concerned with equity and the distribution of research benefits among social groups; and in the late 1980s sustainability concerns were added to the CGIAR objectives. These objectives are not only quite broad but may be in conflict. For example, research that contributes to agricultural growth by improving mechanization might also exacerbate rural poverty (at least in the short run) by reducing farm employment.

Because agricultural research has multiple, sometimes conflicting, objectives some argue that agricultural research can best be evaluated in terms of its overall contribution to social welfare, expressed in economic terms. In fact, economists sometimes argue that economics provides the only valid methods for evaluating agricultural research and guiding decision making.

Research inputs

By "inputs" the CIPP model refers not to human, physical and financial resources, but to the plans and procedural designs on which a research program is based. Agricultural research processes have seldom been systematically studied, and little is known about how to plan effective research programs. In most agricultural research organizations there has been surprisingly little planning until recently.

In the past, research management was highly intuitive; organizations generally attempted to get good researchers, provide them with the resources they needed and let them work in peace. It was taken for granted that research was a "good thing." In such a context, research plans were usually "sales documents" to obtain funds or satisfy accountability requirements, not careful designs that managers and researchers expected to follow.

However, as funding gets tighter and public sector reform processes progress, research organizations are expected to plan and manage their activities to ensure appropriate use of public funds and reasonable returns to social expenditures (Alston, Norton & Pardey, 1995).

Research processes

Unpredictability. Research is quite unlike most standard production processes, in that it is exploratory and inherently risky. The results can not be predicted. For this reason, few agricultural research projects actually produce useful results that are widely applied and have a measurable impact on agricultural production, rural incomes, food supplies and prices or the environment. Because research results are so difficult to predict, techniques normally used for planning, monitoring and evaluating standard production processes capital-intensive development projects are less useful for managing agricultural research.

Temporal dimensions. Crops and livestock have defined production cycles that cannot be overcome by simply expanding R&D facilities or budgets. New tools like biotechnology are revolutionizing some R&D processes and allow researchers to shorten some research processes. For example, it takes several years to produce an improved potato variety and then it takes several more years to provide farmers with a significant amount of improved seed. The long time between the initiation of research, the production of useful results and their widespread application on farms complicates evaluation activities and often makes it difficult to attribute changes in agricultural practices to specific research activities.

Spatial dimensions. Agricultural systems are finely tuned to local conditions; and so are their pests, diseases and other production constraints. In order to respond to farmers' needs in different regions, national agricultural research systems (NARS) need to have research facilities throughout the country, with each center implementing a unique set of research activities that responds to local needs and interests. The dispersion and multiplicity of research

activities make agricultural research evaluation more complex than the evaluation of for example centralized industrial research programs.

Research products

As noted earlier, the principal outputs of research are knowledge and information. The latter is typically documented in research reports and publications or is embedded in new technologies. Different types of research have different objectives, outputs, uses and impacts. In one case the goal may be to improve a research process, and the research output is an input into another research process. In another case research may generate a new crop variety. In a third case, research may identify a "new" virus affecting crop yields. In the first case the research result might be knowledge used in future research but never published. In the second case the result is a variety grown by farmers. And in the third case the result is a publication in a peer-reviewed journal. Such outputs are difficult to measure and standardize.

While the ***direct outputs*** of research are things like knowledge, information and technologies, as noted earlier the ***ultimate goals*** are to contribute to agricultural growth, food security, poverty alleviation and environmental sustainability. Clearly, there are many intervening steps between the production of direct research outputs and the achievement of these broad socio-economic goals. Adoption studies and impact assessments attempt to measure the short- and long-term social, economic and environmental effects of agricultural research.

Methodological challenges for evaluators

As the foregoing implies, agricultural research evaluation presents important methodological challenges. It is difficult to measure *research outputs* like information and knowledge in ways that allow comparisons over time or among different programs. In industrial countries, where a high proportion of research results are published, bibliometrics or citation studies can provide useful measures of research output and impact on science. However, this approach is less useful in developing countries where relatively few research results are published and bibliographic data bases are incomplete.

Measuring the use of research outputs and the *impact* on production, incomes, and the environment is much more difficult yet. An important consideration is the substantial time lag that often occurs between research activities, farmer adoption of new technologies, and ultimate impacts on social welfare and the environment. In many cases, the total time lag from the initiation of research to its ultimate impact may be 10 years or more.

With the passage of time, agronomic, economic and social conditions often change dramatically, making it difficult to distinguish the changes due to research from those due to other factors. Research is sometimes naively viewed as the principal source of technological change and agricultural innovation. However, farm-level innovation has many sources, of which agricultural research is seldom the principal one (Biggs, 1990).

As a result of these factors it is generally very difficult to establish clear cause-effect relations between specific research activities and changes in production, food security, social welfare and the environment. Economists employ various statistical techniques to isolate the contribution of research, but the poor quality of data available and the use of "heroic assumptions" often compromises the objectivity and precision of economic calculations, many of which appear more rigorous and scientific than they in fact are (Horton, 1986; 1990).

As an indication of the difficulty and cost of evaluating the impact of agricultural research on the economy, society and the environment, Collinson & Tollens (1994) estimate that if rigorous ex ante and ex post economic evaluations were to be done for all projects of the

international agricultural research centers, it would require a doubling of the entire budget of the CGIAR!

3. Current state of agricultural research evaluation

Agricultural research activities and organizations are frequently evaluated, but experience with agricultural research evaluation is poorly documented, particularly in developing countries. This section is based largely on Daniels (1987), Novoa (1989) and Novoa and Horton (1994) and on my personal observations, particularly in Latin America.

In most cases, it appears that agricultural research evaluation has been initiated in response to external demands for accountability. Agricultural research organizations in industrialized countries have evaluated their research projects, programs and units for many years; and by now various types of evaluation are taken for granted (Nielson, 1992; Ayres, 1993). In most developing countries, experience with evaluation is more recent. Where it has proceeded furthest, evaluation has generally evolved through three phases:

> **Phase 1.** External evaluation of donor-funded projects, with limited participation of local professionals
>
> **Phase 2.** Development of internal evaluation capacity to meet the external accountability requirements of donors and government agencies
>
> **Phase 3.** Establishment of internal evaluation systems that provide information for both external accountability requirements and for internal decision-making

It appears that most agricultural research organizations in Sub-Saharan Africa are still in Phase 1 while most of those in Asia and Latin America are in Phase 2. Systematic internal evaluation procedures, the norm in agricultural research organizations in industrial countries, are found in the more advanced organizations in Asia and Latin America.

Types of evaluation

In agricultural research organizations, many different things are evaluated for different reasons at different points in time. This has led to considerable confusion as to what evaluation is (and isn't) and how to improve it (Daniels, 1987; ISNAR, 1991; Novoa, 1989). Table 2 identifies nine types of evaluation that are commonly carried out in agricultural research organizations, and indicates the evaluation methods commonly associated with them. This typology is based on a review of literature (Horton, Ballantyne &Uribe, 1990) and on personal observations.

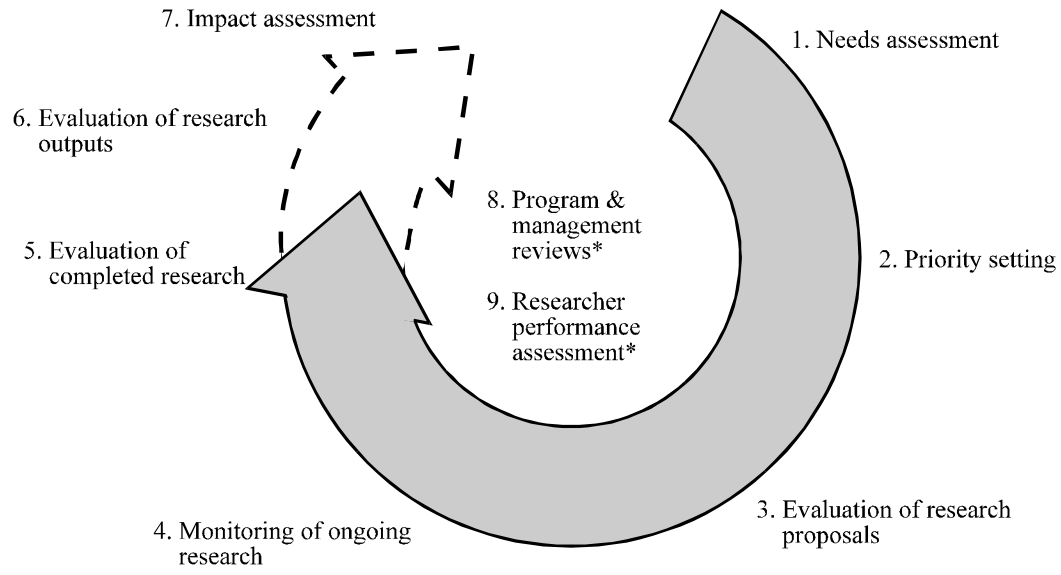
The research project cycle (Gapasin, 1993) provides a useful device for visualizing how the different types of evaluation relate to one another (Figure 3). The first five types refer to evaluations that take place at different moments during the agricultural research process. The sixth and seventh types refer to evaluations of the products and effects of research. The eighth type refers to the evaluation of research units, and the nineth concerns the evaluation of the principal research input, the researcher.

Table 2. Types and methods of agricultural research evaluation

Evaluation type	Evaluation methods
1. Needs assessment	– Expert judgment

	– Literature review – Farmer, market & consumer surveys – Macro-economic studies
2. Priority setting	– Expert judgment – Ex-ante assessment of economic impact – Social or environmental impact assessment – Scoring methods – Participatory planning
3. Review of research proposals	– Peer review – Expert review – Stakeholder review
4. Monitoring of on-going research: – Internal reviews – External reviews	 – Review by peers – Review by external panel of experts
5. Evaluation of completed research	– Project completion reports – External evaluation by group of experts
6. Evaluation of research outputs – Reports and publications – Technologies	 – Peer review – Bibliometrics – On-farm trials – Producer, market & consumer surveys – Participatory evaluation with clients
7. Impact assessment – Adoption studies – Economic impact assessment – Social & environmental impact assessment	 – On-farm surveys & trials – Ex-post economic evaluation – Social & environmental assessment methods
8. Program and management reviews – Internal reviews – External reviews – Internally commissioned external reviews – Audits	 – Review by peers – Review by external panel of experts – Review by external panel of experts – Financial audit – Comprehensive audit
9. Staff performance assessment – Selection – Promotion – Annual evaluation of current performance	 – Committee review for consistency with future job requirements – Same as for selection – Review by supervisor for consistency with current job requirements

Figure 3. Types of evaluation in relation to the agricultural research project cycle



*Program and management reviews and researcher performance assessments are concerned with evaluation of researchers and research units responsible for managing the entire reseach project cycle.

The evaluation of researcher performance is an important component of an evaluation system. However, it is not dealt with here. This paper's focus is on the potential application of program evaluation concepts and methods in agricultural research organizations, and program evaluation applies mainly to the evaluation of an organization's context, inputs, processes and products (evaluation types 1-8). Assessment of researcher performance is a specialized field of its own..

Needs assessment

Agricultural research is generally initiated to solve a problem, and some sort of needs assessment generally precedes the initiation of a research activity. In relatively weak agricultural research organizations that depend on external donor funding, needs assessment is often carried out by external consultants. In more advanced systems, research needs and opportunities are usually assessed by local researchers based on expert judgment and a review of literature. Increasingly, farmers and other intended users of research results are being involved in needs assessment, through diagnostic surveys and participatory stakeholder assessments (Chambers, Pacey & Thrupp, 1989; Scoones & Thompson, 1994).

Evaluation of alternative lines of research (priority setting)

When broad research needs and opportunities are established, the resources available are generally found to be insufficient to permit research on all the important or promising areas. Hence, alternative lines of research need to be evaluated and priorities set for the allocation of scarce resources among the competing areas. As funding declines and demands on agricultural

research organizations proliferate, more and more attention is being paid to “ex-ante” assessment and priority setting.

Economists have perfected rigorous economic methods for setting priorities based on comparisons of estimated future research costs and benefits (Alston, Norton & Pardey, 1995). For it to approximate real-world conditions, ex-ante economic evaluation requires highly skilled analysts, a considerable amount of quantitative data, translation of research costs and benefits into a common economic numeraire and complex computations. In many cases, a number of “heroic” assumptions are used. Such methods are applied in some advanced research organizations such as the Australian Centre for International Agricultural Research (Davis & Lubaleva, 1995) and in some CGIAR centers (Kelly, Ryan & Patel, 1995). But they are generally beyond the reach of agricultural research managers in developing countries, who continue to rely on less rigorous, often highly intuitive, methods.

Simplified approaches, involving "scoring" of alternative research areas in terms of their feasibility and expected contributions to research objectives have been developed and are gaining popularity. Such exercises are often carried out in workshops involving researchers and representatives of various groups of stakeholders (Collion & Gregory, 1993; Collion & Kissi, 1993; CSIRO, 1994; ISNAR, 1995; Franzel, Jaenicke & Janssen, 1996).

Evaluation of research proposals

Research proposals are generally reviewed prior to approval and funding. The most common type of evaluation at this point is review by peers or experts in the field. Peer review is an excellent tool for judging a proposal's scientific merit and quality, but is generally less useful for evaluating the relevance and potential impact of proposed research (Gapasin & Uribe, 1993).

Traditional peer review processes are sometimes modified to incorporate the views of research clients or stakeholders. Forms of "stakeholder review" are the norm where producers and industry groups finance agricultural research activities and sit on review boards for competitive grant schemes. In some cases, the expected contributions of research projects to productivity, equity and resource conservation goals are estimated and review panels consider this information when evaluating research proposals.

Monitoring of ongoing research

Monitoring of on-going research is an important aspect of day-to-day research management. Traditionally, research monitoring has been informal; program leaders have checked up on research activities by interacting directly with scientists and by visiting field trials and laboratories periodically. But increasingly monitoring is being formalized. Where research is externally funded, donors generally call for periodic external reviews of project activities. And in most organizations, researchers are now required to submit periodic progress reports on their work (generally each 6-12 months, or at the end of each crop cycle). Internal reviews, organized at the level of research programs, centers or entire organizations, are increasingly common in both national and international organizations (Goldsworsthy & Gapasin, 1993). For example, each of the international centers affiliated with the CGIAR holds an annual internal program review, to review the past year's research and to establish guidelines for the next year's work. The centers are also encouraged to commission external reviews of their major research areas and their management each few years.

Evaluation of completed research

Where research is managed in projects, with defined objectives, time horizons and budgets, project completion reports are commonly prepared. This is the norm where research is funded

by external donors (Peterson, 1993). However, few agricultural research organizations in developing countries manage their research in well-structured projects (Gapasin, 1993). Instead, researchers have worked with a great deal of freedom, developing lines of research that continue more or less indefinitely. For the lack of a project system, completed research activities are in fact seldom evaluated. This limits documentation of research results as well as learning lessons from different research strategies and processes.

Evaluation of research outputs

The types of evaluation discussed above all occur during the research process itself, and their results may serve as inputs into the direction of an on-going research process (management support). In contrast, the evaluation of outputs and impacts occur after research has been completed. Hence, they serve mainly to document results of previous research (accountability) or to learn lessons from past experience that might be of use in planning and managing future research activities.

Research results are generally documented in reports and publications or they are embodied in new technologies. Research reports and scientific papers are commonly peer reviewed prior to publication, and the published research outputs of programs and units are sometimes evaluated using citation studies or bibliometrics. Bibliometric evaluations are most appropriate in industrial countries where a high proportion of research activities lead to the publication of articles in international professional journals. In many developing countries where few research results are published and bibliographic data bases are poor, they are less valuable (Ballantyne & Uribe, 1993).

New technologies, like varieties, pest management techniques or post-harvest practices, are generally evaluated in experiments on research stations or on farms. However, in traditional research designs, there is seldom a valid comparison between a potential new technology and farmers' current practices. moreover, the traditional evaluation criterion is physical yield or some other agronomic trait, rather than economic performance or social acceptability.

Since about 1970, social and natural scientists have engaged in several types of multidisciplinary research (Janssen & Goldsworthy, 1996). This has often involved systems research and participatory methods for involving farmers in the evaluation of new technologies, at increasingly early stages of the R&D process (Ashby, 1991; Ashby & Sperling, 1995; Horton, 1984; Rhoades 1984; Sperling, Loevinsohn & Ntabomvura, 1993; Tripp, 1991).

Impact assessment

The ultimate purpose of agricultural research is to contribute to agricultural growth, food security, poverty reduction and protection of the environment. Hence, those who sponsor research and carry it out have great interest in the socio-economic and environmental impacts of agricultural research. In agricultural research, impact assessment has been generally considered the domain of economists, and many economic impact studies have been published (Echeverria, 1990). Most of these report high rates of return to investments in agricultural research. Since economic evaluation requires highly trained economists and places heavy demands on quantitative data and analysis, it is beyond the capabilities of many agricultural research organizations in developing countries. However, it is possible to conduct relatively simple economic impact studies that produce useful results. In Canada, an evaluation of R&D accomplishments in the Northern Research Centre revealed that "benefits of just six of 62 rate R&D accomplishments significantly exceeded the entire R&D budget for the ten-year study

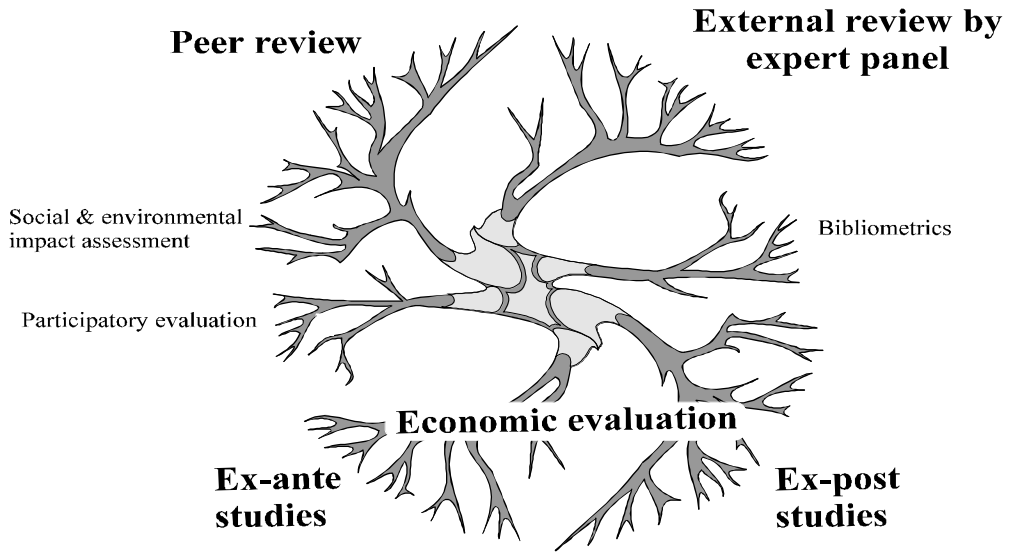
period" (Moore, 1992). An impact study at the International Potato Center produced a similar result (Walker & Crissman, 1996).

Program and management reviews

The various types of evaluation discussed above are used to assess research activities and outputs at different stages of the research and development process. In contrast, institutional reviews are carried out to assess the functioning and performance of research organizations or units (e.g., institutes, departments or programs) as a whole. As governmental agencies seek ways to reduce their budgets, decentralize decision making and ensure high levels of performance, institutional reviews, which look at both programs and management, are increasingly common. Such institutional reviews are now commonplace in the agricultural research organizations of industrial countries and in the CGIAR.

The CGIAR organizes external program and management reviews (EPMRs) of each of its centers approximately every 5 years. In preparation for EPMRs, centers are encouraged to organize Internally-Commissioned External Reviews (ICERs) of their major programs and of management. Initially, most external reviews were little more than peer reviews of center research. However, the focus of external reviews has gradually shifted toward ensuring that appropriate management mechanisms are in place (including planning, monitoring and evaluation procedures) to guarantee adequate levels of effectiveness, efficiency and overall performance (CGIAR, 1995b; Özgediz, 1993; 1995).

Figure 4. Branches of agricultural research evaluation



Development agencies have financed reviews of many national agricultural research organizations and systems as inputs into planning their support for national agricultural

research (Reijmerinck & Uribe 1991). Increasingly, agricultural research organizations are organizing their own reviews. Reviews of entire organizations are generally carried out at major turning points, for example when a new director assumes leadership or the organization's mission or funding is seriously questioned. Some organizations have developed regular procedures for periodic review of their research programs or centers. For example, Brazil's Agricultural Research Corporation (EMBRAPA) and Argentina's National Institute for Agricultural Technology (INTA) review each of their research centers periodically (Borges & Horton, 1994; Hogg, 1994).

Financial auditing is commonplace in agricultural research organizations. Comprehensive audits, which look not only at use of financial resources but at goal attainment, are being experimented with in some cases, like INTA in Argentina.

Branches of agricultural research evaluation

We have seen that many different types of evaluation are carried out in agricultural research organizations. One might expect, then, that many different evaluation methods would be used. This is not the case. Relatively few methods are used, irrespective of the evaluation issues and the types of evaluation called for. As Figure 4 shows, the methods used most frequently fall along three main branches:

- Peer review
- Expert review
- Economic evaluation

Other methods less frequently used fall along three minor branches:

- Participatory evaluation
- Bibliometric studies
- Social and environmental impact assessment

These six branches of agricultural research evaluation are discussed in the following paragraphs.

Peer review

Peer review is applied in virtually all stages of the research process and also to evaluate research products. It is most common in the review of research proposals, internal program reviews and the review of scientific reports and publications.

Many research organizations have formal procedures for reviewing research proposals, reports and publications. The extent to which these procedures are adhered to is not known. Personal experience and observations in several organizations lead me to believe that peer review procedures often become ritualized or subverted. In the first case, they become ineffective because all proposals or products pass the test. In the second case they lose objectivity and serve to enforce personal agendas, rather than as an objective, dependable evaluation procedure.

Expert review

Expert review has much in common with peer review, but has the important difference that while the term "peer" connotes a review by ones equals, "expert" connotes a review by someone more knowledgeable than oneself. For example, in peer review a research proposal is evaluated by colleagues or scientists with similar specializations and levels of expertise. But in "expert review," the review is supposedly done by someone with superior knowledge, perhaps the director of research, or a leader in the field.

One of the most common types of formal evaluation carried out in agricultural research organizations is the external review carried out by a panel of experts. Such panels are generally formed on an ad hoc basis for a short period of time (usually one to three weeks) to evaluate a research project, program or center.

External reviews by experts have their origins in the traditions of peer and expert review within research organizations and in the accountability procedures of development agencies, which generally call for periodic external review of projects during execution and final evaluations upon project completion. (Some would argue that their origins go back further, to the Spanish Inquisition or further still to God's judgment on Adam, Eve & the serpent in the garden of Eden!)

Periodic external review of research programs and management is the norm in many industrial countries and in the CGIAR, and it is beginning to appear in developing countries.

Many external reviews and internally commissioned external reviews conducted by experts are poorly designed and executed. In an analysis of more than 1,000 evaluation reports in ISNAR's library, more than half were found to lack explicit evaluation frameworks or methodologies (Horton, 1992). Most of these were team efforts that were guided by brief, vague terms of reference; the evaluation issues were seldom defined; there was apparently no systematic data gathering or analysis; and reporting left much to be desired.

In such cases, there is little chance of objectivity or replicability, reports are bound to vary widely in form and content from evaluation to evaluation, and the selection of "experts" on the panel is perhaps the single factor which influences the evaluation results. Despite their many weaknesses, external review by expert panel often have major impacts on the future of the unit being evaluated and on the people involved.

Economic evaluation

The most rigorous commonly used methods for evaluating agricultural research are drawn from production economics. What economists refer to as "ex-ante" economic analysis can be used to assess research needs and set priorities; "ex-post" analysis can be used to evaluate the economic impact of completed research. For economists these two types of analysis are generally seen as two sides of the same coin. The main difference is that in ex-ante analysis economists attempt to *predict* future streams of research costs and benefits associated with specific lines of research; whereas in ex-post analysis they attempt to *measure* actual cost and benefit streams. In both cases, research is viewed as an investment and the analysis results in estimated net economic benefits and rates of return.

Economic evaluations tend to be complex and costly, and for this reason, they are often carried out as independent research projects. Many economic evaluations of agricultural research have been done by young economists as dissertation research projects. Many agricultural economists treat evaluation as a sub-discipline within economics and argue that there is no substitute for economic analysis as a guide to resource allocation. The following two statements illustrate this point:

> "The two primary purposes of research evaluation from the point of view of participants in the CGIAR are (a) to obtain and justify funding support and (b) to support making decisions about the allocation of funds among alternatives…This leads us to focus this paper on the quantitative approaches to evaluation, based on economic principles. Other approaches are better suited to other types of decisions, such as day-to-day management of scientists. They may serve a useful monitoring role or provide

> data that can be used in a quantitative assessment, but they cannot substitute for the methods discussed here" (Alston & Pardey, 1995).
>
> "In meeting research users' needs, however, there appears to be no substitute for the use of economic analysis, preferably as an **economic surplus model** (i.e., a representation of changed consumer and producer surpluses induced by adoption of research findings) in its comprehensive or simplified forms, to assist in priority setting among programs and subprograms" (World Bank, 1996).

Economic evaluations are generally highly rigorous and their results are valued in policy discussions and efforts to justify funding for agricultural research. However, the impact of economic evaluations on decision making is sometimes called into question. Averch (1991) notes that two streams of ex-post agricultural research evaluation have developed in the United States, and that:

> "Neither stream appears to influence resource allocation decisions, although this is not something that can be held against research evaluation. All such decisions involve factors other than rates of return, especially in highly charged political environments" (Averch, 1991).

Participatory evaluation

Turning now to the minor branches of agricultural research evaluation, beginning in the 1970s, a number of anthropologists and sociologists have worked with agronomists in developing participatory research and evaluation methods, including farming systems research and participatory rural assessment (PRA). Development agencies have funded a considerable amount of participatory on-farm work, and non-governmental agencies are heavily involved in PRA (Scoones & Thompson, 1994; Farrington & Bebbington, 1991; Pretty & Chambers, 1993; Okali, Sumberg & Farrington, 1994). Participatory needs assessment and evaluation exercises have produced many important insights into research and innovation processes. However, they have seldom become institutionalized in national agricultural research organizations (Merrill-Sands et al, 1991).

Bibliometrics

Bibliometric or scientometric studies are frequently used to evaluate research and the impact of research units on scientific literature in industrial countries. However, bibliometric studies are rare in developing countries. This is partly due to the highly applied nature of much agricultural research in developing countries and the limited publication of their research results there or elsewhere. The lack of comprehensive data bases containing research publications is another limitation in developing countries. Despite these limitations, bibliometrics can be expected to be used as an evaluation tool more extensively in the future, as databases are improved and appropriate computer hardware and software become more readily available.

Social and environmental impact assessment

In the past, when agricultural research was focused more sharply on increasing agricultural productivity, there were few rigorous assessments of the social or environmental impacts of agricultural research. However, as public pressurus have mounted for agricultural research organizations to address issues of social welfare and environmental sustainability more directly, interest in social and environmental impact assessment has increased significantly. A number of agricultural research organizations have begun relevant studies, and more can be expected in the future.

Disciplinary roots of agricultural research evaluation

Why are peer review and expert review so prevalent in agricultural research organizations? Why is cost-benefit practically the only systematic method used? And why have broader program evaluation theory and expertise been virtually ignored? The answers appear to stem fundamentally from the "hard-science" culture of agricultural research organizations and the roles played by professionals trained in different scientific disciplines in them.

Mix of scientific disciplines in agricultural research organizations

The professional staff members of agricultural research organizations are principally natural scientists specialized in fields such as agronomy, genetics, plant pathology, and biotechnology. Until the last few years, most agricultural research organizations were run by natural scientists, most of whom started as bench or field scientists and gradually worked up the career ladder to become department heads, program leaders and ultimately directors of entire research organizations.

Many scientists now working in international centers and NAROs in developing countries received post-graduate degrees in colleges of agriculture of land-grant universities in the USA. At many of these institutions, there has been a long tradition of multi-disciplinary research involving agronomists and agricultural economists. The economists often estimated production costs and analyzed the costs and benefits of alternative technologies. These tasks were useful for translating results of agronomic trials into farmer recommendations. Through this work, agricultural economists perfected production economics and many econometric techniques.

In contrast, very few social scientists have worked in a multi-disciplinary mode with natural scientists in a university setting. The few social scientists who worked in colleges of agriculture were generally involved in extension, communication and home economics -- areas that were considered to be less prestigious than research per se.

Over time, international agricultural research centers have employed more agricultural economists. Initially, economists were practically "forced" into the international centers, by donors like the Ford Foundation, which felt that economic analysis was an essential complement to the core research work. Economists' cost-benefit work was perceived as useful for developing farmer recommendations. Later, economists were called upon to estimate the economic benefits and rates of return to successful research efforts. Again, donors often demanded this work. Results of impact assessments were useful for demonstrating that agricultural research was a good investment. More recently, many economists are focusing on ex-ante impact assessments, which can guide the allocation of scarce resources among alternative research areas.

In the 1980s and 1990s, as the funding crisis set in and management issues became more complex, agricultural economists have moved into key management positions in the CGIAR. Presently, all the top managers of the CGIAR (the Chairperson of the CGIAR, the Executive Secretary of the CGIAR Secretariat and Chairperson of TAC) are economists, as are the directors of five of the 16 international centers. These individuals clearly have an influence over the types of evaluation that are carried out in CGIAR centers and the methods used.

In sharp contrast to agricultural economists, few social scientists have ever been employed in agricultural research organizations and fewer yet have reached key management positions. In the 1970s and 1980s, several social scientists were posted to CGIAR centers through special programs like the Rockefeller Foundation's Post-Graduate Fellowship Program. These young professionals usually became involved in multi-disciplinary research, and many of them made

significant contributions to the development of client-oriented research approaches. However, after their initial postings, relatively few social scientists stayed on in the centers as researchers. The centers often viewed social science research to be of marginal interest, and the social scientists seldom found attractive prospects for research careers in the centers.

In the same period, some international centers hired educators and communicators to develop training, publication and information programs. Many of these programs blossomed in the 1980s, but, with the severe reductions in funding and staffing since 1990, they have been cut back disproportionately, to protect resources for the "core" research activities of the centers. In many cases, social scientists' contracts have been terminated. Their programs have sometimes been closed. In other cases they have been sharply scaled down and leadership has been transferred to natural scientists. As a consequence, the absolute and relative numbers of social scientists working in international agricultural research centers has fallen in recent years.

Relatively few agricultural economists and extremely few social scientists are employed by national agricultural research organizations in developing countries. Several economists, sociologists and anthropologists have been engaged, along with agronomists, in farming systems research and participatory technology development efforts funded by development agencies. After their zenith in the 1980s, when these projects ended, the social scientists usually departed. Core-funded positions were seldom available, and even where they are, the salary levels were seldom adequate to retain good economists. Few managers considered the it important to keep a social scientist on the staff, unless perhaps to prepare extension bulletins or write speeches. Two decades ago, it was estimated that only one sociologist was employed in agricultural research organizations for each 100,000 research scientists (van Dusseldorp, 1977). The ratio does not seem to have changed much since then.

Links between the disciplines and evaluation methods

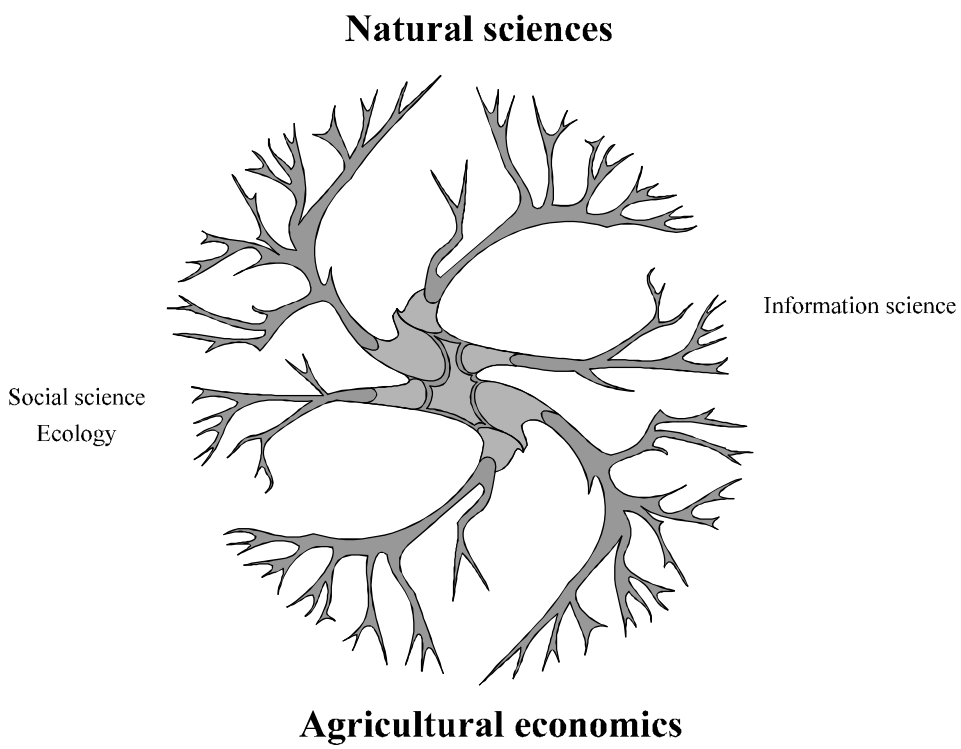
There is a clear link between the dominant disciplines and the main branches of evaluation in agricultural research organizations. When top management is called on to evaluate research activities or programs, the logical reaction is to do the type of evaluation one is familiar with and to draw on in-house expertise. Consequently the prevailing mix of disciplines in agricultural research organizations has a major influence on the evaluation methods used.

For the reasons outlined above, the methods used to evaluate agricultural research have their disciplinary roots mainly in the natural sciences and agricultural economics. Some less frequently used methods have their roots in the social science, information science, and other disciplines (Figure 5).

The natural sciences. Peer review and expert judgment have long traditions in the natural sciences; hence it is not surprising that these are the most generally accepted and widely used methods for evaluating agricultural research. A few economists and management specialists have contributed to development of procedures for external reviews by expert panels, but expert judgment remains the prevailing standards for evaluation in these exercises.

Agricultural economics. When managers have needed more complex evaluation studies, they have generally turned to the agricultural economists on their staff. And predictably, the agricultural economists have used economic methods. Natural scientists have provided information inputs (for example, estimates of the probability of success in particular areas of research), but these studies remain well rooted in the discipline of agricultural economics.

Figure 5. Disciplinary roots of agricultural research evaluation



The social sciences and other disciplines. The relatively minor branch of participatory evaluation has disciplinary roots in anthropology and sociology. Agricultural scientists, "user-friendly" economists and research clients have also participated in this work. Interestingly, this is the only branch of agricultural research evaluation where economists and social scientists have collaborated to any significant extent. There was considerable activity in this branch of evaluation in international centers and in public-sector agricultural research organizations in the 1970s and 1980s, when client-oriented on-farm research was in vogue. More recently, non-governmental organizations have been particularly active in this area.

The small bibliometrics branch has stemmed mainly from the work of information specialists. In some cases, statisticians, economists or natural scientists have collaborated.

The small branch referred to as "social and environmental impact assessment" has distinct sub-branches which have their disciplinary roots in the social sciences, economics and environmental studies. This is the youngest branch of agricultural research evaluation, and it is presently one of the most vigorous branches. Unlike the other branches, there is as yet no dominant methodology in environmental impact assessment.

5. Prospects for the future

Presently, there are strong external pressures on agricultural research organizations to improve management and accountability, and managers express interest in developing and strengthening their evaluation systems. The donors, development banks and national groups

that fund agricultural research are all demanding evidence that their money is being well spent and that research is producing benefits for society. Agricultural research organizations, along with other public service providers, are expected to employ mechanisms to ensure accountability, transparency and overall performance. These issues are by no means academic, they bear on the very survival of public-sector organizations. Many agricultural research organizations are being significantly "down-sized" and some are disappearing altogether, at least as we know them at present.

The present scenario of increased public scrutiny and demands for improved accountability and performance have led to attempts to introduce program evaluation in some agricultural research organizations. In this section, I review some recent experiences with program evaluation in the CGIAR and discuss some implications of this experience for future efforts with program evaluation in agricultural research.

Recent experiences with program evaluation in the CGIAR

The first attempt to introduce program evaluation to the CGIAR was in 1991, when a symposium on assessing the impact of international agricultural research for sustainable development was organized by the Cornell International Institute for Food, Agriculture and Development (Lee, Kearl & Uphoff, 1992). Lee Sechrest, Robert Boruch and William Trochim were invited to represent the program evaluation community and to speak on the last afternoon of the three-day event. In earlier sessions, presentations were made by evaluators (mainly agricultural economists) from CGIAR centers and organizations associated with the CG system. Unfortunately, there was little interaction between the CGIAR participants and the program evaluators. Some key participants (including the TAC Chair person) left at lunch time just before the program evaluators' presentations were scheduled to be delivered. And there was no follow up after the symposium.

The second opportunity to introduce program evaluation to the CGIAR came a few years later, as a result of the CGIAR's funding crisis and a revitalization program launched by the CGIAR Chairperson, Ismael Serageldin. (Mr. Serageldin is the World Bank's Vice President for Environmentally Sustainable Development.)

In February, 1995, a Ministerial-level meeting was called in Lucerne, Switzerland to discuss a renewal program for the CGIAR with key donors. The resulting *Lucerne Declaration* called for strengthening the assessment of CGIAR performance and impact (CGIAR, 1995a).

In response to the *Lucerne Declaration*, a special workshop was held, in conjunction with a meeting of the CGIAR in Nairobi, Kenya in May, 1995. The workshop was chaired by an eminent agricultural economist, Robert Herdt, who had served a number of years in the International Rice Research Institute and the International Food Policy Institute and heads the agriculture department of the Rockefeller Foundation. The meeting provided a unique opportunity for interaction among key CGIAR stakeholders on the topic of evaluation, because so many donor representatives, directors of CGIAR centers and the centers' evaluation practitioners attended. The exchange of views was quite animated and useful. However, there was no mechanism or method for comparing and contrasting these views, identifying areas of consensus and divergence or reaching a higher-level synthesis or priorities for action. Moreover, there was no attempt to document the presentations or the discussion, and no proceedings were prepared. The principal tangible result of the workshop was a brief presentation to the CGIAR a few days later on a proposed evaluation function.

Later in the year, the CGIAR established the Impact Assessment and Evaluation Group (IAEG) with the following aim:

> "to assist in the development of evaluation as an integral component of all CGIAR activities so that evaluation becomes a part of the CGIAR culture at the researcher and project level, at the center level, and across the system" (CGIAR, 1996).

Chairperson Serageldin has been a prime mover in creation of the IAEG, which has two purposes. The prime focus is "to assess the impact of CGIAR supported activities on poverty, food security and environmental sustainability." In addition, improved evaluation is also expected to "identify ways in which all aspects of CGIAR operations….can be improved" (CGIAR, 1996).

The Group, established in late 1995, has three part-time members. It is chaired by James Peacock, an eminent agricultural scientist who is Chief of the Division of Plant Industries at Australia's Commonwealth Scientific and Industrial Research Organization. The second member is Eugenia Muchnik, Director of the graduate program in agricultural economics at the Catholic University in Chile; previously Ms. Muchnik served on the Technical Advisory Committee of the CGIAR. Initially, the third IAEG member was Eleonor Chelimsky, past President of the American Evaluation Association and founding Director of Program Evaluation and Methodology in the General Accounting Office of the US government. After serving for a year Ms. Chelimsky left the IAEG to dedicate more time to her other evaluation-related activities. At the time of writing, a replacement had not yet been named for Ms. Chelimsky.

One of the first activities of the Group was to organize a workshop on impact assessment and evaluation, in April 1996. This event brought together evaluation practitioners from CGIAR centers (mainly economists), leading program evaluators (mainly psychologists from the USA) and representatives of a number of donors and international organizations close to the CGIAR, like the World Bank.

After the April meeting, the IAEG developed a work plan, issued a contact list for program evaluators around the world and initiated some evaluation studies. These include a meta-evaluation of impact studies already done in CGIAR centers, an economic study of the impact of genetic research and case studies of adoption of center-related technologies. These and planned future activities involve interactions among evaluators at the centers and professional program evaluators.

This initiative is important and promising. However, there are some reasons for concern. Several of the CGIAR participants in the April 1996 workshop felt that the program evaluators invited to represent the evaluation community were poorly informed about the work of the CGIAR and the large amount of evaluation work that had been carried out within the CGIAR to date. Some participants also felt that the program evaluators had relatively little to contribute to evaluation methods in the CGIAR. The fact that the IAEG members serve on a part-time basis and reside on three different continents makes it difficult for them to address the evaluation needs of the CGIAR with persistence and vigor. On a number of opportunities, Center Directors and economists have questioned the adequacy of the IAEG mechanism. Ms. Chelimsky's resignation leaves the Group, at least temporarily, without senior program evaluation expertise.

Conditions for successful introduction of program evaluation

The CGIAR experience and other recent experiments with program evaluation in Latin America (to be presented in a future issue of *Knowledge and Policy*) indicate that in addition to good intentions and even high-level commitment, there are at least three other necessary conditions for successfully introducing program evaluation:

- Effective communication between agricultural researchers, managers and program evaluators
- Adequate professional and financial resources for program evaluation
- Existence of a program evaluation function and a defined role for program evaluators
- A mindset that is receptive to the admission of conceptual frameworks and operating procedures from a wide variety of disciplines

Effective communication

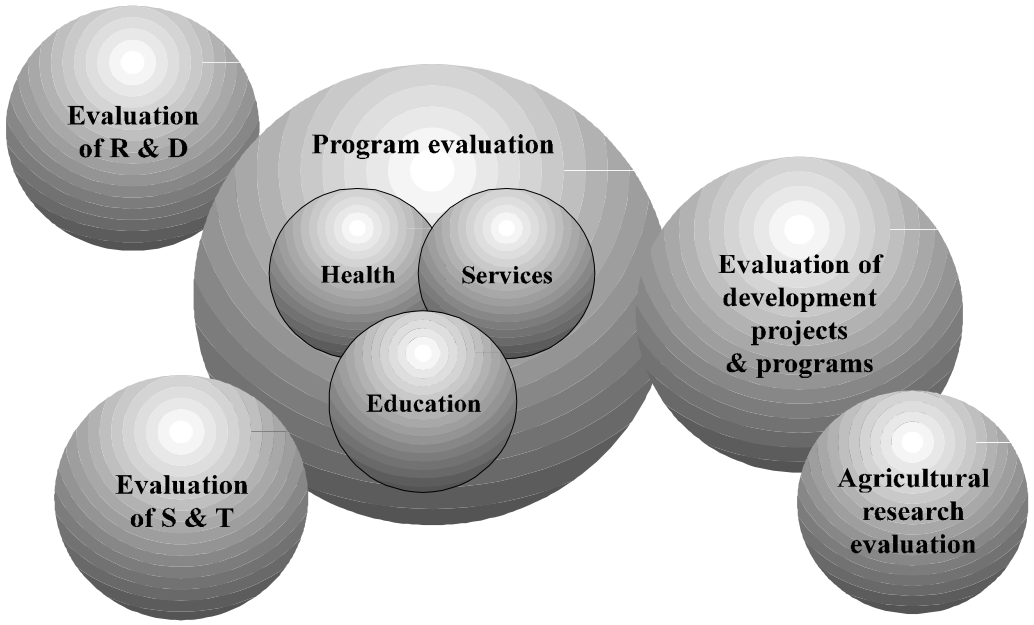
Evaluators who are rooted in different disciplines (e.g., psychology, education, sociology, economics) and who work in different sectors (e.g., health, education, science and technology, agricultural research) tend to operate in specialized spheres, each with its own professional and substantive concerns, value systems, discourse and patterns of communication (Figure 6).

When evaluators in one sphere come into contact with their counterparts in different spheres, they frequently discover that many principles are general and that practical experiences and lessons can be exchanged profitably. This occurs frequently in the "core" spheres of program evaluation. The commonality of disciplinary backgrounds, subject matters, concerns and values allows fruitful communication and the gradual development of evaluation as a new "transdiscipline," with principles that apply across the various source disciplines and sectors. In this sense, Shadish, Cook and Leviton (1991) refer to the foundations of program evaluation as "theories of practice."

There are other cases, however, where spheres of evaluation barely come into contact or where their root disciplines and sectoral issues have been so insulated one from the other that there is little common ground.

In the agricultural research establishment, very few evaluation practitioners have had much contact with the program evaluation community. As pointed out previously, most evaluators in agricultural research organizations are either natural scientists involved occasionally in peer or expert review exercises or agricultural economists who carry out economic studies of various sorts, including economic evaluations. Natural scientists have little professional interest in program evaluation, as their professional careers and their participation in evaluations depend more on their personal contacts and their credibility as subject-matter specialists than on the quality of the evaluations they produce (as quality is defined, for example, in the Program Evaluation Standards). Economists on the other hand find little of interest in program evaluation because they are highly specialized in economic methods, and their professional affiliation tends to be restricted to disciplinary bodies like agricultural economics associations.

Figure 6. Spheres of evaluation knowledge and expertise



S&T - Public-sector science and technology programs
R&D - Private-sector research and development projects

Adequate resources

Program evaluation requires expertise, it takes time and it costs money. Introduction and systematic application of program evaluation in agricultural research organizations requires resources both for methodology development and for design and consolidation of the internal evaluation systems.

In the 1970s, the "Golden Era" of program evaluation in the U.S.A., federally funded social programs were expanding rapidly, billions of dollars were being spent on them each year and millions of dollars were spent on evaluation. According to Shadish, Cook and Leviton (1991) a 1982 U.S. Government Accounting Office survey of federal evaluative activities (excluding Defense and funds for audits and routine management information) reported 228 evaluation organizations employing about 1,400 professionals and spending about $180 million in 1980. Federal funding for evaluation declined in the 1980s but resources are still provided from many sources, and evaluation continues to progress and consolidate as a professional field in the U.S.A.. Evaluation funding and development has followed a similar course in Canada. Government-funded evaluation activity began later in Europe. It is now in a rapid phase of development to respond to demands for evaluation from both national governments and the European Union.

In contrast to the above-mentioned situations where governmental programs were expanding and resources for evaluation were abundant, funding for agricultural research is now declining and resources for evaluation are scarce. As the total funding available to both international and

national agricultural research organizations declines, a manager can spend more on evaluation only by spending less on research per se.

Funding cuts are threatening the very existence of publicly funded research facilities. No manager wants to go down in history as the one who closed the research center. So when faced with the decision of whether to expand evaluation or maintain core research activities, most managers understandably choose the latter.

One important barrier to acceptance of the IAEG and introduction of program evaluation in CGIAR centers is that the IAEG has very few resources that it can offer centers to defray the additional expenses that would be entailed and to motivate them to come on board. The Group also possesses very limited and narrow program evaluation expertise. It has engaged program evaluators through consultancy contracts on a part-time basis, for specific tasks. But the investment of professional resources has been very small relative to the size and evaluation needs of the CGIAR.

A recognized program evaluation function

In the core spheres of program evaluation (health evaluation, educational evaluation and the like), evaluation is professionalized, and practitioners view themselves as "evaluators"; they participate in professional evaluation associations and read and publish in specialized evaluation journals. In contrast, most evaluation practitioners in other spheres, like development project evaluation and agricultural research evaluation are more strongly identified with their base disciplines and sectorial specializations than with evaluation as a profession.

In international agencies, the professionalization of evaluation is hampered by the common practice of rotating staff members among different organizational units, including the evaluation unit. Consequently, few individuals in development agencies become specialized in evaluation or view themselves as professional evaluators. (One exception to this practice is the World Bank, where staff members often remain a number of years in the Operations Evaluation Department.)

National and international agricultural research organizations seldom have specialized evaluation units or staff members dedicated exclusively, or even a large proportion of their time, to evaluation. Most evaluation is done by economists or other scientists on a part-time, ad hoc basis. Hence, those who do evaluation in agricultural research seldom view themselves as evaluators.

For example, CIP's Social Science Department and CIMMYT's Economics Program are responsible for impact assessment in these centers. In the Kenyan Agricultural Research Institute, responsibilities for monitoring and evaluation and for establishment of an institute-wide management information system have fallen to the Socio-Economics Unit. However, the staff members of all these units view themselves as researchers and pursue research careers. Therefore, it is unlikely that internal evaluation can be institutionalized in these units.

Unless agricultural research organizations create specialized evaluation functions and the corresponding structural units, it seems unlikely that program evaluation can be successfully introduced and institutionalized.

A receptive mindset

Promoting knowledge of the field of program evaluation, providing resources or raising the technical standards of individual evaluation studies may be beneficial in and of themselves, but

the most appropriate starting point for developing the practice of program evaluation in agricultural research organizations may be at a more *personal* level.

Returning to the original definition of program evaluation offered at the beginning of this paper, it is clear that program evaluation is carried out first and foremost for the benefit *people* – so that they can make judgments, improvements and informed decisions. This key point has been referred to by Patton (1997) as "fostering intended use by intended users". People will have to become aware and convinced of the use of information generated by evaluation studies *for their own professional use* before program evaluation will become a respected, institutional feature of their organizations.

The suggested starting point is therefore one that puts people in the foreground and assists them in articulating their own, individual, professional needs for evaluations and evaluative information; to express these needs and to determine how similar or different they are from those of their colleagues and other stakeholders. The result would be a shared understanding of the various stakeholders' requirements for evaluative information and respect for the professional needs from which these differences arise.

As indicated above, meetings have already been held in the CGIAR to exchange information and explore evaluation needs and interests. However, they have not been effective in reaching these objectives. As outlined by Guba and Lincoln (1989), an effective exchange and interpretation of divergent views that moves toward a higher-level synthesis of them all requires that all parties have a willingness to

- Work from a position of openness and integrity
- Share power
- Change, if they find the negotiations persuasive
- Reconsider their value positions as appropriate
- Make the commitments of time and energy that may be required in the process

These conditions have generally not been fulfilled in CGIAR meetings where program evaluation has been discussed. Instead, many participants have come with fixed ideas and personal or institutional agendas, and they have left with their ideas and agendas very much intact. More careful preparation, design and facilitation would be required to make real progress in achieving a broader understanding of evaluation needs and opportunities in the CGIAR and a more receptive mindset toward program evaluation.

A useful parallel activity could be for key stakeholder groups to sponsor a comprehensive review and analysis of recent evaluation studies in order to inform themselves and determine the extent to which the information contained in existing evaluations was useful, and the methods employed were credible and convincing. The "metaevaluation" currently being sponsored by the IAEG is a step in the right direction. However, it would be useful if a group of key evaluation CGIAR stakeholders could be more directly involved in the assessment of usefulness and credibility.

From these two activities, the major players in agricultural research would develop a shared understanding from which a mutually useful discourse could emerge. The discourse would encourage the development of a mental set that would encourage the development of evaluation as a "transdiscipline" that would serve them appropriately and well.

6. Conclusion

Evaluation research is about providing useful, appropriate, credible and timely information to people who are faced with decisions. Decisions are seldom taken by one group of decision makers in isolation and they seldom affect a single set of individuals. Therefore, they must satisfy or at least be transparent to all of the stakeholders affected. Information for such transparent and cross-cutting decisions must come, not from narrow-based, self-serving, restricted studies, but from studies that have been undertaken with people in mind – people from a range of levels or areas within a given sector.

Program evaluation, with its roots, as it has shown to be in this paper, in a wide variety of scientific and social fields, and branching as it does into virtually every area of human professional endeavor, could be of enormous assistance in creating a common platform to unite disparate groups in achieving their common goals.

References

Alston, J.M. & Pardey, P.G. (1995). *Research Evaluation Requirements in the CGIAR.* Repo9rt comissioned by the impact assessment task force for the public awareness and resource mobilization committee (PARC) of the CGIAR. Processed.

Alston, J.M., Norton, G.W. & Pardey, P.G. (1995). *Science under Scarcity*. Cornell University Press, Ithaca.

Anderson, J.R., Herdt, R.W. & Scobie, G.M. (1988). *Science and Food: The CGIAR and its Partners.* World Bank, Washington, D.C.

Ashby, J.A. (1991). Adopters and Adapters: The Participation of Farmers in On-Farm Research, In Tripp, R. (Ed.). *Planned Change in Farming Systems: Progress in on-farm Research.* John Wiley & Sons Ltd, Chichester.

Ashby, J.A. & Sperling, L. (1995). Institutionalizing Participatory, Client-Driven Research and Technology Development. *Development and Change*, 26(4).

Averch, H. (1991). *The Practice of Research Evaluation in the United States.* Research Evaluation, 1 (3).

Ayres, C. (1993). Planning, monitoring and evaluation in the research branch of Agriculture Canada. ISNAR discussion paper No. 93-08. International Service of National Agriculture Research (ISNAR), The Hague.

Ballantyne, P. & Uribe, B. (1993). Bibliometrics, In Horton, D., Ballantyne, P., Peterson, W., Uribe, B., Gapasin, D, & Sheridan, K. *Monitoring and Evaluating Agricultural Research: A Sourcebook.* CAB International, Wallingford.

Biggs, S.D. (1990). A Multiple Source of Innovation Model of Agricultural Research and Technology Promotion. *World Development*, 18(11), 1481-1499.

Borges, J., Debela, S., Lusthaus, C., Mackay, R., & Smutylo, T. (1996). *ISNAR Achievements, Impact and Constraints 1991-1996: Final Report.* (processed) International Service for National Agricultural Research (ISNAR), The Hague.

Borges, J.E. & Horton, D. (1994). Planificación, Seguimiento y Evaluación en la EMBRAPA, Brasil, In Novoa, A.R., & Horton, D. (Eds.). *Administración de la Investigación Agropecuaria*. Tercer Mundo Editores, Santafé de Bogotá.

Brady, N.C. (1994). Consultative Group on International Agricultural Research, In *Enclyclopedia of Agricultural Science*, Volume I, Academic Press, Orlando.

Brinkerhoff, R., Brethower, D., Hluchyj, T., and Nowakowski, J. (1986). *Program Evaluation: A Practitioner's Guide for Trainers and Educators.* Kluweer-Nijhoff, Boston.

CGIAR (1995a). *Renewal of the CGIAR: Sustainable Agriculture for Food Security in Developing Countries*, Consultative Group on International Agricultural Research (CGIAR), Washington D.C.

CGIAR (1995b). *Review Processes in the CGIAR: Terms of Reference and Guidelines for External Program and Management Reviews of CGIAR Centers.* CGIAR Mid-term meeting. Document No. MTM/95/11, Consultative Group on International Agricultural Research (CGIAR), Washington D.C.

CGIAR (1996). Impact Assessment and Evaluation Group (IAEG) report to the CGIAR. CGIAR Mid-term meeting. Document no. MTM/96/12, Consultative Group on International Agricultural Research (CGIAR), Washington D.C.

Chambers, R., Pacey, A. & Thrupp, L.A. (1989). *Farmer First: Farmer Innovation and Agricultural Research*. Intermediate Technology Publications, London.

Collinson, M.P. & Tollens E. (1994). *The Impact of the International Agricultural Research Centers: Measurement, Quantification, and Interpretation.* Issues in Agriculture, No. 6, Consultative Group on International Agricultural Research (CGIAR), Washington, D.C.

Collion, M.H. & Gregory P. (1993). *Priority Setting at CIP: An Indicative Framework for Resource Allocation.* International Potato Center (CIP), Lima.

Collion, M.H. & Kissi A. (1993). Learning by doing: developing a programme planning method in Morocco. *Public Administration and Development,* 13(3), 261-270.

CSIRO (1994). *Setting Research Priorities in CSIRO*. Commonwealth Scientific and Industrial Research Organization (CSIRO), Camberra.

Daniels, D.(Ed.). (1987). *Evaluation in National Agricultural Research*. International Development Research Centre (IDRC), Ottawa.

Davis, J. & Lubulwa, G. (1995). *Integration of Research Evaluation Analysis into Research Institution Decision-Making: An Overview of Progress at ACIAR.* Economic Evaluation Unit. Working Paper No. 17. Australian Centre for International Agricultural Research, Canberra.

Echeverría, R.G. (1990). Assessing the Impact of Agricultural Research, In Echeverría R.G. (Ed.), *Methods for Diagnosing Research System Constraints and assessing the Impact of Agricultural Research*. Vol. 2, International Service for National Agricultural Research (ISNAR), The Hague.

Farrington, J. & Bebbington, A. (1991). *Institutionalisation of Farming System Development.* Paper for FAO Consultation on the Institutionalisation of Farming Systems Development, Overseas Development Institute, London.

Franzel, S., Jaenicke, H. & Janssen, W. (1996). *Choosing the Right Trees.* International Service of National Agriculture Research (ISNAR), The Hague.

Gapasin, D. (1993). Research Project Management, In Horton, D., Ballantyne, P., Peterson, W., Uribe, B., Gapasin, D, & Sheridan, K. *Monitoring and Evaluating Agricultural Research. A Sourcebook.* CAB International, Wallingford.

Gapasin, D.P, & Uribe, B. (1993). Peer Review, In Horton, D., Ballantyne, P., Peterson, W., Uribe, B., Gapasin, D, & Sheridan, K. *Monitoring and Evaluating Agricultural Research. A Sourcebook.* CAB International, Wallingford.

Goldsworthy, P. & Gapasin, D.P. (1993). Internal Program Review, In Horton, D., Ballantyne, P., Peterson, W., Uribe, B., Gapasin, D, & Sheridan, K. *Monitoring and Evaluating Agricultural Research. A Sourcebook.* CAB International, Wallingford.

Guba, E. & Lincoln, Y. (1989). *Fourth Generation Evaluation*. Sage Publications, Newbury Park.

Hogg, D. (1994). Planificación, Seguimiento y Evaluación en el INTA, Argentina, In Novoa, A.R., & Horton, D. (Eds.). *Administración de la Investigación Agropecuaria*. Tercer Mundo Editores, Santafé de Bogotá.

Horton, D. (1984). *Social Scientist in Agricultural Research: Lessons from Mantaro Valley Project, Peru.* International Development Research Centre (IDRC), Ottawa.

Horton, D. (1986). Assessing the impact of international agricultural research and development programs. *World Development*, 14 (4), 453-468.

Horton, D. (1990). Assessing the Impact of International Research: Concepts and Challenges, In Echeverría, R. *Methods for diagnosing research system constraints and assessing the impact of agricultural research*. Vol.2. International Service for National Agricultural Research (ISNAR), The Hague.

Horton, D., Ballantyne, P.& Uribe, B. (1990). *The Many Faces of Monitoring and Evaluation a Progress Report on the M&E Literature.* ISNAR Staff Note 90-98. International Service for National Agricultural Research (ISNAR), The Hague.

Horton, D. (1992). Two Institutional Aspects of Impact Assessment: Evaluation of, and in, National Agricultural Research System, In Lee, D.R. (Ed.) *Assessing the Impact of International Agricultural Research for Sustainable Development: Proceedings from a Symposium at Cornell University.* CIIFAD, Ithaca.

ISNAR (1991). *Highlights of a Consultation on Monitoring and Evaluation of Agricultural Research*. International Service for National Agricultural Research (ISNAR), The Hague.

ISNAR (1995). *Research Program Formulation.* Training Module, International Service for National Agricultural Research (ISNAR), The Hague.

Janssen, W. & Goldsworthy, P. (1996). Multidisciplinary Research for Natural Resources Management: Conceptual and Practical Implications. *Agricultural Systems,* 51, 259-279.

Joint Committee on Standards for Educational Evaluation (1994). *The Program Evaluation Standards.* SAGE Publications, Thousand Oaks.

Kelly, T.G., Ryan, J.G. & Patel (1995). Applied Participatory Priority Setting in International Agricultural Research: Making Trade-offs Transparent and Explicit. *Agricultural Systems*, 49, 177-216.

Lee, D., Kearl, S. & Uphoff, N. Assessing the impact of international agricultural research for sustainable development. Proceedings from a symposium at Cornell University. Cornell International Institute for Food, Agriculture and Development. Ithaca.

Love, A.J. (1991). *Internal Evaluation. Building Organizations from Within.* SAGE Publications, Newbury Park.

Mackay, R.(1996). Undertaking ESL/EFL programme review for accountability and improvement. In T. Hedge & N. Whitney (Eds.). *Power, Pedagogy, and Practice*. Oxford: Oxford University Press.

Merrill-Sands, D.M., Biggs, S.D., Bingen, R.J., Ewell, P.T., McAllister, J.L., & Poats, S.V. (1991). Institutional Considerations in Strengthening on-Farm Client-Oriented Research in National Agricultural Research Systems: Lessons from a Nine-Country Study. *Experimental Agriculture*, 27(33), 342-373.

McLean, D. (1988). *The Logical Framework in Research Planning and Evaluation.* ISNAR working paper 12. International Service for National Agriculture Research (ISNAR), The Hague.

Moore, W. & Newstead, R. (1992). Evaluation of Research and Development Accomplishments: Northern Forestry Centre. *The Canadian Journal of Program Evaluation,* 7(1), 41-51.

Nielson, J. (1992). Planning, Monitoring and Evaluating in the Agricultural Research Center, Washington State University. ISNAR discussion paper No. 92-05. International Service for National Agriculture Research (ISNAR), The Hague.

Novoa, A.R.(Ed.). (1989). *Agricultura, Tecnología y Desarrollo: Como se Evalua la Investigacion Agraria en America Latina.* Tercer Mundo Editores, Santafé de Bogotá.

Novoa, A.R., & Horton, D. (1994). Planificación, Seguimiento y Evaluación de la Investigación agropecuaria en las Americas: Una síntesis. In Novoa, A.R. & Horton, D.E. (Eds.). *Administración de la Investigación Agropecuaria.* Tercer Mundo Editores, Santafé de Bogotá.

Okali, C., Sumberg, J., & Farrington, J. (1994). *Farmer Participatory Research.* Intermediate Technology Publications, London.

Özgediz, S. (1993). Organization and Management of the CGIAR system: a review, *Public Administration and Development,* 13(3), 217-231.

Özgediz, S.(1995). *Strengthening Evaluation in the CGIAR: Needs and Options.* Draft, Consultative Group on International Agricultural Research (CGIAR), Washington D.C.

Pardey, P., Roseboom, J.& Anderson, J. (1991). Regional Perspectives on National Agricultural Research. In Pardey, P., Roseboom, J. & Anderson, J. (Eds.). *Agricultural Research Policy: International Quantitative Perspectives.* Cambridge University Press, Cambridge.

Patton, M. Q. (1997). Utilization-Focused Evaluation: The New Century Text. Sage Publications, Thousand Oaks.

Peterson, W. (1993). Development-Agency M&E, In Horton, D., Ballantyne, P., Peterson, W., Uribe, B., Gapasin, D, & Sheridan, K. *Monitoring and Evaluating Agricultural Research. A Sourcebook.* CAB International, Wallingford.

Pretty, J.N. & Chambers, R. (1993). *Towards a Learning Paradigm: New Professionalism and Institutions for Agriculture,* Discussion paper 334, Institute of Development Studies, Sussex.

Reijmerinck, J.& Uribe, B. (1991). *ISNAR Diagnostic Reviews: An Analysis of Recommendations.* International Service for National Agricultural Research (ISNAR), The Hague.

Rhoades, R.E. (1984). *Breaking New Ground.* International Potato Center (CIP), Lima.

Ruttan, V. (1982). *Agricultural Research Policy.* Minneapolis: University of Minessota Press.

Sartorius, R. (1996). *The Third Generation Logical Framework Approach: Dynamic Management for Agricultural Research Projects.* European Journal of Agricultural Education and Extension, 2(4).

Scoones, I., & Thompson, J.(1994). *Beyond Farmer First.* Intermediate Technology Publications Ltd, London.

Scriven, M. (1993). Hard-won lessons in program evaluation. *New Directions for Program Evaluation*, 58.

Shadish, W.R.(Jr.), Cook, T.D. & Leviton, L.C. (1991). *Foundations of Program Evaluation.* SAGE Publications, Newbury Park.

Sperling, L., Loevinshohn, M. & Ntabomvura, B. (1993). Rethinking the Farmer's Role in Plant Breeding: Local Bean Experts and On-Station Selection in Rwanda. *Expl. Agric.*, 29, 509-519.

Stufflebeam, D.L. (1984). The CIPP model for program evaluation, In Madaus, G.F., Scriven, M. & Stufflebeam, D.L. (eds), *Evaluation models: Viewpoints on educational and human services evaluation*. Kluwer, Boston.

Tribe, D. (1994). *Feeding and Greening the World.* CAB International, Wallingford.

Tripp, R. (Ed.). (1991). *Planned Change in Farming Systems: Progress in on-farm Research.* John Wiley & Sons Ltd, Chichester.

van Dusseldorp, D.B.W.M. (1977). Some Thoughts on the Role of Social Sciences in the Agricultural Research Centres in Developing Countries. *Netherlands Journal of Agricultural Sciences*, 25.

Walker, T. & Crissman, C. 1996. Case studies of the economic impact of CIP-related technologies. Lima:CIP. 157p

World Bank, Operations Evaluation Department (1996). *Achievements and Problems in Development of National Agricultural Research Systems.* Report No. 15828. The World Bank, Washington, D.C.