ASSESSING THE PAST TO PLAN THE FUTURE

A paper presented at the
Regional Workshop on Research Program Evaluation

Dhaka, Bangladesh, 1 - 3 October, 1984

by

Josette Murphy, PhD
Senior Research Officer

International Service for National Agricultural Research
Oranje Buitensingel 6
The Hague
The Netherlands

1. INTRODUCTION

Government officials and development leaders often have a negative view of evaluations. The very word calls to mind a short visit by a group of outsiders who disrupt regular activities, take a quick trip along good roads, and come back to tell you everything you did wrong. Evaluations are seen as focusing on the past, with the purpose of inspecting an on-going or completed activity, and possibly auditing the use of funds.

This misconception is rooted in the fact that evaluations in developing countries have usually been organized and used by donors in relation to project assistance, with little involvement of national officials. It is not surprising, therefore, that for many research leaders and managers, evaluations have a negative image and are not seen as an integral part of good management.

In fact, evaluations can have their main focus on the past, the present, and the future, and they form powerful tools for national planners, policy makers and research managers themselves, outside of any donor-funded project. Evaluations should become an integral part of effective planning, whether it is planning for all research activities in a country, a specific scientific program, or the activities of a particular institute. Evaluations enable a government or a research institute to build lessons from past experience into revised priorities and the preparation of future programs. They also enable a program leader to adjust the program during implementation and maybe revise its objectives so that research results are a real contribution. Few developing countries have much experience with this innovative use of evaluations, even though more are now becoming aware of its potential.

2. PURPOSE OF THE WORKSHOP

The purpose of this workshop is to illustrate the potential of evaluations for Bangladesh research, and to encourage the participants to integrate internal evaluations in their management practices.

Three key points will be emphasized throughout this workshop:

1. Arrangements for evaluation procedures should be made when a research program is being designed, not as an afterthought.

2. Evaluations should not be seen as isolated activities but should build upon regular monitoring procedures and involve staff from both inside and outside the program.

3. Evaluation reports should be used for planning future activities as well as to assess current achievements, and to justify future investments.

The three presentations this morning and the two technical sessions will deal more specifically with the evaluations of research programs, defined as a set of experiments and research activities which are implemented to reach a clearly identified objective(s). I use the term program in its scientific, not administrative sense. A program does not necessarily coincide with a research institute or with a department, it is likely to involve researchers from several disciplines, several departments, sometimes even in different institutes. A program usually includes several projects.

In my presentation, I will first present some general principles, introduce the several possible uses of program evaluations, and discuss the importance of setting up clear objectives and measurable indicators of achievement at program design. I will finally introduce the type of information which is likely to be included in the scope of work for a program evaluation.

My colleagues will expand on some of the points I am introducing, Dr. Baird on the diverse uses of evaluations and Dr. Nestel on the information required. This will be followed this afternoon by a Technical Session in small groups, to draw a list of information which should be covered in the scope of work of a program evaluation in Bangladesh.

In the second Technical Session tomorrow at 10:15, you will define the objectives of a sample program and select indicators of achievements for it.

Dr. Mook tomorrow morning will step back from today's focus on individual programs to discuss manpower and training assessment, because this is a type of evaluation which is essential to successful research activities.

During the last plenary session tomorrow afternoon, each group will report the scope of work it will have prepared this afternoon, and a general discussion will follow.

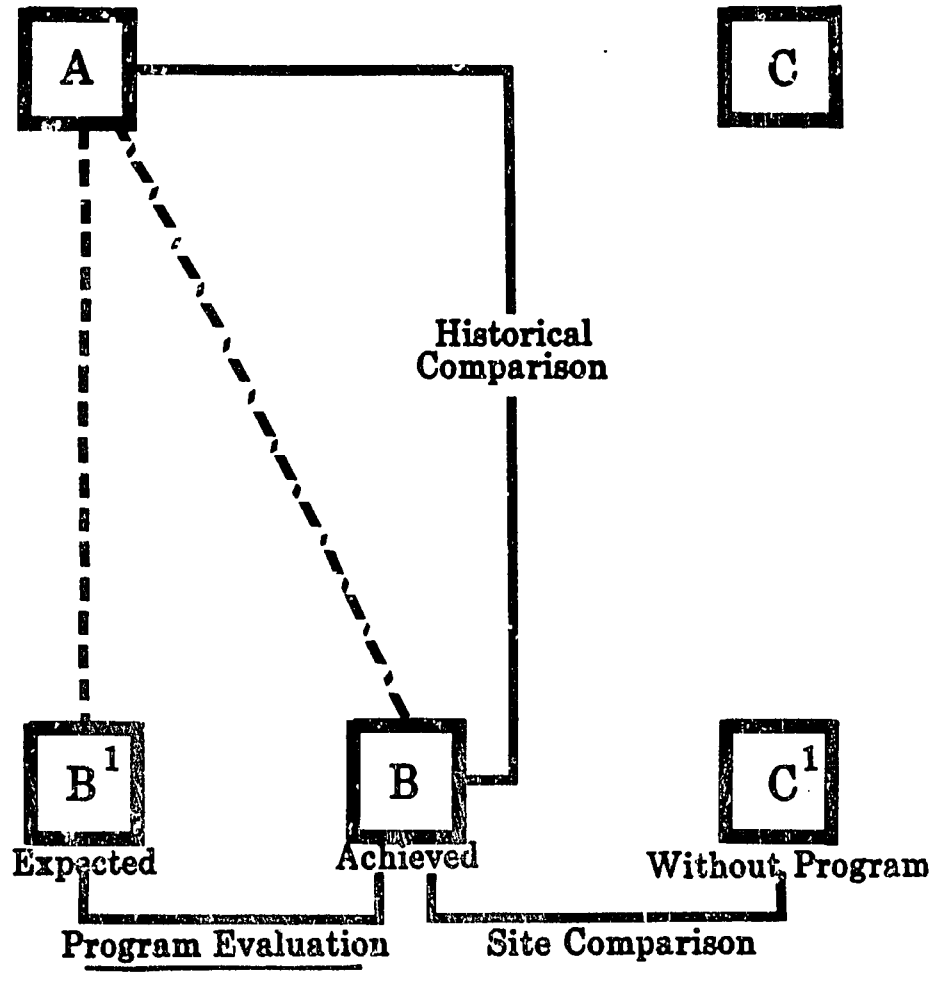
3. GENERAL PRINCIPLES FOR EVALUATIONS

Allow me to begin with some general principles on evaluations:

1. An evaluation begins at program design, by setting up clear, specific, verifiable indicators of achievements.
   An evaluation always entails a relative judgement: you can only evaluate a situation by comparing it with another one, therefore there must first be agreement on what will be acceptable as a standard indicator of achievement. There are several possibilities (diagram 1):

   - compare the situation at time B with what it was at time A.
   - compare the situation at time B with that in C1, where no program existed.
   - compare the situation at time B with an expected B1 which was defined when the program was designed in the first place.

Diagram 1: An evaluation is a comparison



The third comparison is the only valid evaluation of a program.

The first one is a historical study. The second may sound like a control, and appeal to scientists. However, in the real world, there is no such thing as a controlled situation because one cannot keep other factors constant. This site comparison, like the historical study, has its use but neither permit a judgement of a program's achievements. Therefore the first rule in evaluations is that an evaluation begins when a program is being designed, by setting up clear, specific, verifiable indicators of achievements for that program and by specifying how the achievements will be measured. This absolute requirement has two immediate benefits:

- It forces the program designers to clearly express what the objectives of the program are and what results are expected, in very concrete terms.

- It requires specifying how progress and achievements will be measured and therefore establishes the basis for monitoring procedures. This leads us to the second key principle:

2. There can be no valid evaluation without adequate mechanisms for monitoring, record-keeping and reporting throughout the life of the program.

   In order to compare the results achieved against those which were expected, data and information must be available for the evaluation team to understand what has actually been achieved to date and what occured during implementation. It is time-consuming and difficult to retrace this type of information afterwards, and sometimes simply impossible. It is much easier to keep track of program implementation as it occurs, through systematic record-keeping and reporting.
   However, this internal data is not sufficient, because of the third principle:

3. No program functions in isolation, and therefore an evaluation must place the program in the institutional, political, social and economic context in which it is implemented.

   Judgement on the quality of program design and implementation cannot be passed by just looking at whether the expected results were attained. It should seek to understand why the achieved results are as they are, always differentiating between factors internal and external to the program.

   We all could give many examples of programs which were technically sound and competently implemented but could not achieve the expected research results because of extraneous factors, such as a change in government priorities or a cut in funding. Should an evaluation find that expected results were not achieved, it will be essential to trace whether this was due to constraints beyond the control of the researchers. Therefore the evaluation team will need information on the context in which the program was implemented.

4. Finally, an evaluation of a research program must very clearly differentiate between achievements of the program's research results and the contribution of the program to a broader development objective. Both types of evaluations are valid and in fact complementary, but they are not interchangeable.

   An evaluation limited specifically to one program covers internal factors (program design, resources, implementation ....), those external factors which influence resources and implementation, and the scientific validity and potential of research results. The results it evaluates are those directly derived from the program: a new or improved technology in most cases.

   An evaluation of the contribution of the program to broader development objectives will cover the same factors plus two other sets: the fit of the program in the overall research and development plan of the country, and actual adoption of research results and its impact on production, income, or whatever the development objectives were. This introduces numerous non-research factors of regional, national and even international dimensions. It is also likely that other research programs become relevant factors in understanding the contribution of one program to development.

   Two conclusions arise from this: first, an evaluation of the contribution of a research program to development is more complex than a regular evaluation of program implementation, it considers a broader scope of topics, and requires a different cluster of expertise to conduct. Its results have also more diverse use for general planning and reassessment of priorities, not only for research but also for development services.

   Second, an evaluation of the developmental impact of a cluster of research programs will be greatly facilitated if evaluations of each individual program are first conducted, but only if these limited evaluations are conducted with compatible procedures. This does not mean that the same team must evaluate each individual program, nor that identical scope of work have to be used, only that coordination at the preparatory stage is essential. This was a prime consideration when organizing this workshop: the general scope of work you will be preparing will make such compatible evaluations possible if desired.

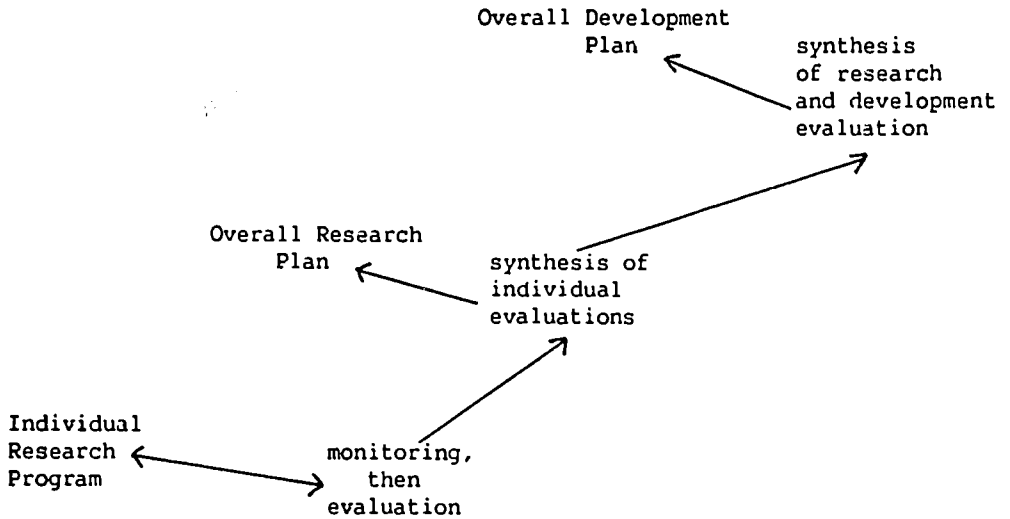
4. RAPID REVIEW OF DIFFERENT PURPOSES AND SCOPES OF EVALUATIONS

Three types of purposes are possible for program evaluations:

1. Adjusting a program while it is still possible to do so.

2. Understanding the factors which influenced - positively or negatively - the achievements of expected program results, and using these lessons from experience in future programming.

3. Understanding the research and non-research factors which influenced actual contribution of a technology to development, and therefore draw lessons for research and for development policy and services.



The purpose of monitoring an on-going program is first to confirm that things are being implemented as planned. It makes it possible to notice bottlenecks and problems before they have caused too much damage.

The purpose of evaluating a completed program - or a particular phase of a program - is to find out whether it led to the expected results. It may also seek to explain what, in the way the program was designed and implemented, facilitated or hampered reaching the desired result. It is therefore necessary, as with monitoring, that the program be clearly identifiable, with well defined expected results against which actual results can be measured.

The evaluation of a completed program can also seek to show what contribution the results achieved made to development activities and to the people who were expected to benefit from them. In this case, the evaluation goes beyond the program as originally planned to see if the reasoning underlying the decision to implement it was correct, and it must also consider the performance of other development activities and services.

Dr. Baird will discuss these purposes in the context of Bangladesh. Here I will only emphasize that they are not necessarily mutually exclusive: an evaluation of a completed activity may be conducted strictly to determine that it has been satisfactorily completed and that the bottom line, so to speak, can be drawn. It can also be conducted to help decide what further activities should be proposed and how their design and implementation plans should follow or differ from what had been accomplished in the completed program.

Furthermore, it is important to keep in mind that research leaders and politicians in a country need more than a pile of evaluation reports for individual programs. They require a synthesis of all existing and possible research achievements available to the country. In other words, they need to understand the capacity and the potential of the entire research system in the country in order to approve an overall research plan which uses the system to its full efficiency to contribute to national development priorities.

A comprehensive system review is more likely to be necessary at an early phase in the development of an overall research plan, when on-going activities do not yet form a coherent set of programs (Diagram 2). More mature institutes should have developed mechanisms to link programs to priorities and to periodically reassess their main objectives, but history has shown that a long existence is no guarantee of maturity. An overall review could also become necessary even in mature research systems should the government revise its development priorities, or should a drastic change occur in research capacity, technology potential, world economy or some other factors which modify agricultural potential and therefore research requirements.

One can well imagine a two-tier system in the case of a country such as Bangladesh, which has a number of mature research institutes as well as a coordinating body involved in the identification of research priorities and overall planning. There could be agreement among the institutes and the coordinating body so that monitoring and evaluation procedures used for individual programs are compatible (but not necessarily identical). This would greatly facilitate a comparative analysis and synthesis of these evaluations, and would provide the coordinating body with an extremely solid basis for planning. This is being developed for contract research, it could be extended to all research activities.

It would also reinforce the position of research when dealing with government authorities, because it would provide a clear picture of research currently underway, of resources used, and of results achieved. Too often, the only type of evidence that research leaders can present to policy makers to justify further investment in research are evidences at development level (production). This is valid, and indeed numerous studies have shown very high rates of return to investment in research. It is not very useful to research leaders in the case of programs which have not yet reached completion but need support to continue, or when the potential impact of a technology on development is being blurred through constraints outside research, such as unfavorable marketing conditions or extreme weather conditions.

5. INCORPORATING ACHIEVEMENTS STANDARDS INTO PROGRAM DESIGN.

I will now discuss the importance of systematically planning for an evaluation when designing a research program. This requires clear definitions of the program objectives, of its expected results, and of the indicators which will make it possible to confirm whether the results have been achieved. It can be summarized unto this type of table (table 1). Most of you will recognize it right away as being very similar to the Logical Framework that the United States Agency for International Development uses in its project papers. One should not assume from its origin that the logical framework (or logframe) is only of interest for donors or as part of an outside-funded project. The logframe is simply a systematic configuration of logical reasoning that everyone of us uses implicitely whenever we make a plan of action, be it a research program or a personal decision such as planning a trip.

Diagram 2

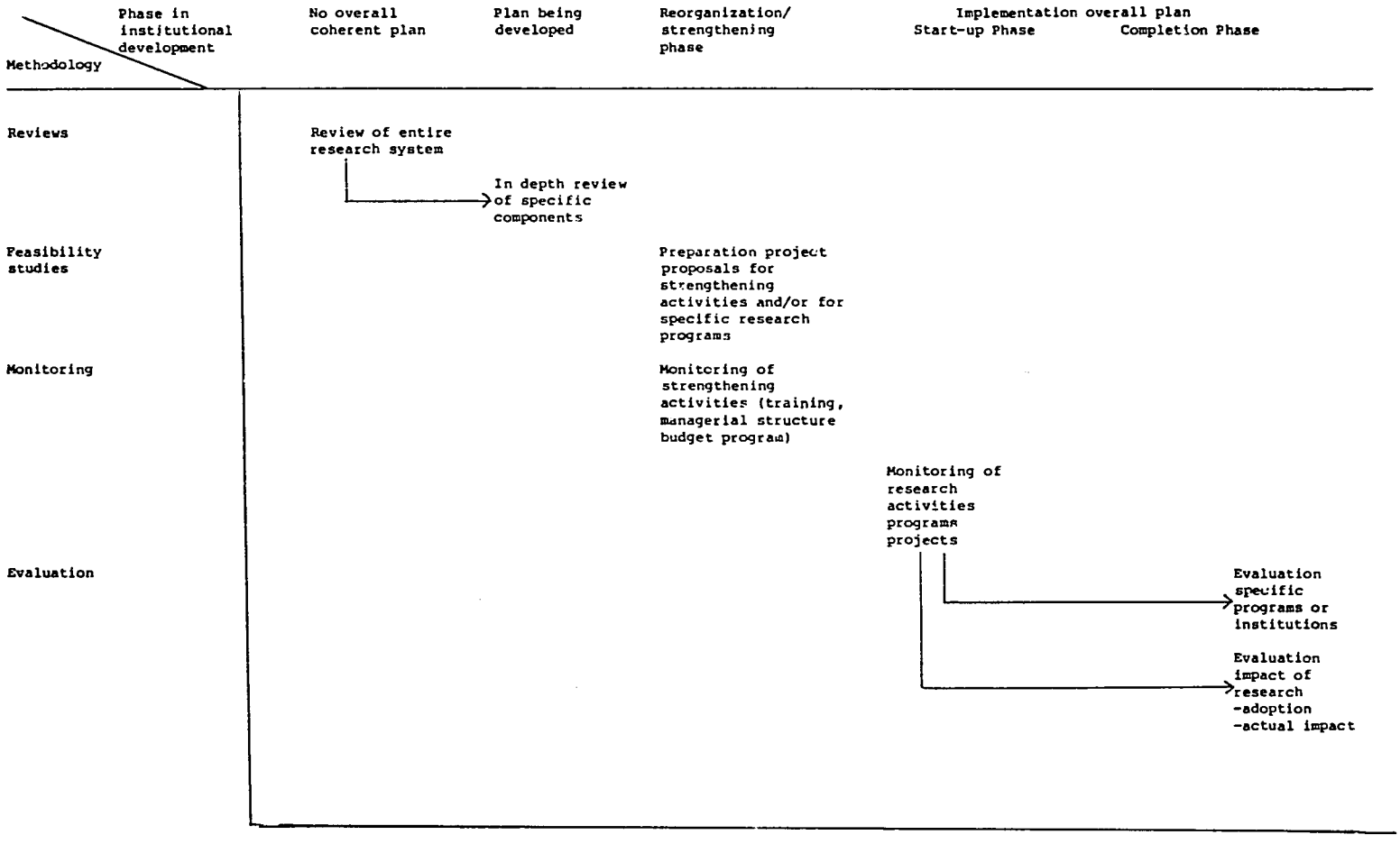


Diagram 2 Phasing between phases in institutional development and types of evaluative activities

Table 1: Logical Framework

	Narrative Summary	Verifiable indicators	Means of verification	Important assumptions
Then ↑	GOAL contribution to overall economic development goals	processing and marketing labor opportunity income	national statistics	- political stability - no drastic environmental changes - no change in world prices
If Then ↑	PURPOSE contribution to agricultural development purposes	increased production or better efficiency	- adoption rate - farm surveys -	- necessary services available - economic environment favorable - no drastic environmental changes
If Then ↑	OUTPUT new or improved product (variety) or technology	specifications on desired characteristics	- laboratory and stations records - certification - on-farm testing results	- continuous support (budget, staff) - underlying scientific reasoning correct - no drastic environmental changes
If	INPUT Actions to be taken under the program	- work plan - schedule of activities	- quarterly reports - accounting and other administrative records	- budget disbursed as planned - staff available

The logframe is a four by four matrix which helps organize the various levels of objectives of an activity and sets up some related parameters to the achievement of each level. I will review the four rows from bottom to top:

Inputs: activities undertaken under the program, with the expectation that implementing those inputs will lead to the production of the outputs.

Outputs: those achievements (variety, technology, knowledge) which derive from the inputs and are not dependent upon other activities.

Purpose: a desired agricultural development scenario for which the research output is necessary but not always sufficient.

Goal: in the broader context of national development, a desired economic achievement for which the agricultural development purpose is necessary but not always sufficient.

Each of these levels of objectives is defined in the particular context of a program in the first column "narrative summary". For example, a breeding program (input) may be implemented to provide a new high yielding variety (output) which would lead to increased production (purpose) which in turn would make it possible to reduce import (goal).
The second and third columns "Verifiable indicators" and "Means of Verification" specify what type of evidence could be taken as sign of achievement of each level of objectives, and how that evidence could be found and measured. The last column, too often taken for granted in development activities, lists those factors not controlled by the program but which influence its implementation and chances for success. For example, changes in world prices of a commodity could influence the purpose to goal relation.

An evaluation of a research program would use primarily the output and purpose narratives, and the two central columns (objectively verifiable indicators, and means of verification). If the program is correctly implemented, then the information necessary to calculate whether the selected indicators were fulfilled (the means of verification) will be gathered routinely throughout the course of the program, and analysed at regular intervals to satisfy reporting requirements.

This is not sufficient however, because a good evaluation should interpret its findings and analyze what caused delays or discrepancies in reaching expected outputs and purposes. In other words, an evaluation does not just fill in the central cells but analyzes the cause and effect relations between cells.

It is essential to give much thoughts to the selection of indicators and the means of verifications. Indicators are not always quantifiable, but they must very clearly measure a causal relationship between the two levels they measure, inputs to outputs, or outputs to purpose, and they should not be dependent on other inputs or outputs.

Even if it is not quantitative, an indicator should be very explicit and as precise as possible, and objectively measurable - "certification of a better wheat variety" is not a valid indicator: the concept of "better" is always relative. If a program's objective is to breed a variety of wheat which fits in a given cropping pattern, and yields more than the traditional one, then an appropriate indicator may be certification of a variety with planting date in November, which matures in less than 150 days, and which consistently yields more than 2 tons/ha in real farm conditions. Means of verification in this case would be records from the certification boards, records from trials in experimental conditions, and results of on-farm testing and verification. Different evaluators should come to the same results when giving a value to an indicator. They may have different opinions as to why actual results match or do not match the pre-established indicators.

6. PREPARING A SCOPE OF WORK

I would like now to narrow my presentation down to one particular aspect of the preparation of an evaluation: deciding what type of information it is both necessary and sufficient to cover. Let us be very clear that there can never be a standard scope of work valid for any research program. This afternoon you will discuss what type of information is necessary when evaluating a program in Bangladesh, but even the general scope of work which you will prepare would have to be adapted to the particulars of any specific program.

Remember the general principle introduced earlier on the importance of placing the program in context. The scope of work for an evaluation will cover several main topics, some dealing directly with various aspects of the program, some dealing with the context in which it takes place, and some dealing with changes the program is expected to bring about.

In the course of its services to national agricultural research systems and organizations, ISNAR has identified nine topics which are likely to require consideration in an evaluation and should therefore be covered in its scope of work. The topics are as follows:

A - The country setting
B - Structure, organization and place of the research program
C - Planning and budget
D - Human resources
E - Facilities, equipment and supplies
F - Scientific activities and achievements
G - Management of the program
H - Communication linkages
I - Contribution of the research program to development.

The division among these areas is in great part a matter of convenience when gathering information, and they do overlap. Management in particular is not really a separate area but is part of all the others. What really matters is that the areas needing coverage go beyond the implementation of research itself. They cover elements from the national situation in which research results will be applied, and specifically include the policy environment which influences program implementation and adoption of research results.

I will now introduce each topical area and discuss its relative importance, which varies with the purpose and intended audience of the evaluation.

A. COUNTRY SETTING

Some may be surprised to see this first area of Country Setting included in a program evaluation, especially for evaluations conducted by a national team. Yet many aspects of a research program can be assessed only in relation to the situation of the overall research system in the country and to the conditions and potentials of its agricultural sector.

How broad a coverage of the country setting is needed in an evaluation report depends in part upon its purpose, particularly whether it will be used in communications with foreign colleagues or donors, or with national policy-makers. Any document prepared for general release needs more information on the country setting than an internal document does. However, the information is necessary to the evaluation even when it is entirely internal.

B. STRUCTURE, ORGANIZATION AND PLACE OF THE RESEARCH PROGRAM

Two distinct issues are addressed in this area: (1) internal structure and organization -- how the program is organized, how it is internally set up to perform its functions; and (2) contextual structure and organization -- how the program fits into the national agricultural research system as a whole, and how it interacts with other components of the system.

This area can be very brief for an internal evaluation of a program implemented within one institute, it may require detailed attention from the evaluation team in the case of a joint activity involving staff from several institutes. Questions of lines of authority, division of labor, timing of various projects are then more complex and can become crucial factors in the success of the program.

C. PLANNING AND BUDGET

Again this section can be brief: basically limited to a review of the program objectives and program budget, if the purpose of the evaluation is simply to find out whether the program is being implemented as planned. However, if the purpose of the evaluation is broader and includes a reassessment of the program objectives, then the team will need to understand how the objectives were identified in the first place, and how they fit within the overall research and development priorities of the country.

To assess whether the program budget is adequate, the team needs to compare this budget with two requirements: on one hand, wit the work and staff required to implement the program with reasonable efficiency, and on the other hand with the overall resources available for research.

D. HUMAN RESOURCES

Manpower is the very heart of any operation. Well trained, dedicated and productive staff can make all the difference between an effective program and an ineffective one. For this reason, information on the number of staff with various levels of education, experience, and training is collected.

As with the budget, data on manpower allocated to a program should be analyzed in the context of the overall size and level of research personnel in the system.

In an evaluation of an on-going program, the team will also wish to consider eventual training requirements for the remaining life of the program. If insufficient training is found to have been a constraint in a completed program, this is a key lesson for future planning.

E. FACILITIES AND EQUIPMENT

Information on the number, size, and condition of facilities and equipment should be related to that on manpower and training, because equipment is of little use in the hands of staff who are not adequately trained. This area also covers the location of experimental stations and on-farm activities, an essential point with important lessons for future programs and possibly for planning a reorganization of research infrastructure in a country.

I also wish to emphasize that this area covers more than scientific and experimental apparatus. We included in the checklist questions on the adequacy of library facilities and access to world-wide information.

F. SCIENTIFIC ACTIVITIES AND ACHIEVEMENTS

This area is the one which first comes to mind when one organizes an evaluation of a program: it covers what research activities have been conducted, how, and with what results. It is in many ways as much an evaluation of the researchers as of the research, and can be conducted only by individuals who are themselves competent researchers.

The evaluators will wish to review the work plans and experimental designs for each activity under the program. Tracing the life history of a few protocols selected at random can be very enlightening: it will highlight strengths and weaknesses at every level in the implementation of the program, from the criteria taken in consideration by the researcher when designing the protocol, to how the work was actually conducted, results interpreted and reported.

The evaluation will also identify research results already achieved, comparing them with the indicators of achievements specified in the program, and assessing the extent to which those achievements have been recognized outside the program.

G. MANAGEMENT

In a broad sense, management encompasses almost all areas of a program's operations. A separate list is given as a matter of convenience, but it should be used in conjunction with every other list as appropriate. The particular emphasis centers on where responsibility and influence is placed, and the extent to which those with responsibility have reasonable control or influence over the resources and conditions with which they are expected to operate.

H. COMMUNICATION LINKAGES

Agricultural development and increased productivity often depend on a country's ability to form and manage effective two-way relationships between technology developers and technology users. The success of a research program also depends on the effective operation of linkage with other research organizations within and outside the country, development agencies, and policy makers.

Communication linkages can be official, with formal mechanisms to ensure exchange of information, such as regular meetings, lines of reporting, or official visits. These are fairly easy to describe. However, the existence of formal communication mechanisms on paper does not mean that any communication actually takes place. We all know of committees which have an official existence but never met.

The importance of informal mechanisms - communications between individuals rather than between positions - varies from country to country. They are sometimes the most effective form of communication. This is why the list for this area includes questions on the nature of the means of communication, on how actively those means are used, and on how effective they are.

I. CONTRIBUTION OF THE RESEARCH PROGRAM TO DEVELOPMENT

This area does not duplicate area F on scientific activities and achievements but builds up upon it to find out how the scientific outputs of the program are contributing to development. By definition, an evaluation can assess contribution only if scientific results have already been achieved and if sufficient time elapsed for a contribution to be possible.

An evaluation as this level, often called an impact evaluation, cannot be limited to tracing the adoption of research results and (hopefully) subsequent increase in production. When assessing any change in production, it is never correct to place the praise or blame only on research activities. A change in the agricultural sector or in the national economy is always the result of interactions between many agro-ecological, technical, social, economic, institutional and policies variables, of which research is only one. This means that the scope of work of a program evaluation at this level must include many questions beyond research activities. It will have to deal with extension and other services, rural infrastructure, marketing, processing, actual adoption rates, and impact on production and possibly on income, labor, and nutrition. The evaluation becomes more complex and time consuming, but its usefulness goes well beyond a simple assessment of whether a program was implemented as planned.

7. CONCLUSION

In this presentation, I have highlighted some general principles in evaluation methods, discussed the areas likely to be covered in a scope of work, and introduced the various ways in which evaluation findings can be used. From research output to agricultural purpose to national goal, research program evaluations form a useful tool for research leaders and policy-makers. In the next presentation, Dr. Guy Baird will discuss in more depth the various ways in which evaluation findings can be used.