PN-ABC-585

READINGS IN RESEARCH PROGRAM EVALUATION

A background document prepared for the
Regional Workshop on Research Program Evaluation

Dhaka, Bangladesh, October 1 - 3, 1984

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

The International Service for National Agricultural Research (ISNAR) began operating at its headquarters in The Hague, Netherlands on September 1, 1980. It was established by the Consultative Group on International Agricultural Research (CGIAR), on the basis of recommendations from an international task force, for the purpose of assisting governments of developing countries to strengthen their agricultural research. It is a non-profit autonomous agency, international in character, and non-political in management, staffing and operations.

Of the thirteen centers in the CGIAR network, ISNAR is the only one which focuses primarily on national agricultural research issues. It provides advice to governments, upon request, on organization, planning, manpower development, staff requirements, financial and infrastructure requirements, and related matters, thus complementing the activities of other assistance agencies. Additionally, ISNAR has an active training and communications program which cooperates with national agricultural research programs in developing countries.

ISNAR also plays an active role in assisting these national programs to establish links with both the international agricultural research centers and donors.

ISNAR is supported by a number of the members of CGIAR, an informal group of approximately 30 donors; it includes countries, development banks, international organizations, and foundations.

READINGS IN RESEARCH PROGRAM EVALUATION

A background document prepared for the
Regional Workshop on Research Program Evaluation

Dhaka, Bangladesh, October 1 - 3, 1984

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

Table of Content

Preface

		Page
I	Checklists for preparing a scope of work	1
	- The country setting	3
	- Structure, organization and place of the research program	11
	- Planning and budget	13
	- Human Resources	15
	- Facilities, equipment and supplies	17
	- Scientific activities and achievements	19
	- Management of the program	20
	- Communication linkages	22
	- Contribution of the research program to devel	23
II	Example for a scope of work	24
III	Incorporating Indicators of achievements into program design	28
IV	Suggestions for planning an evaluation	34
	1. Clear evaluation objectives	34
	2. Definite but flexible plans	34
	3. Efficient use of existing information	36
	4. Purposeful field visits	36
	5. Team preparation	37
	6. Quality control	38
V	Guidelines for non-experimental data collection	40
	1. Observations	40
	2. Interviews	42
	3. Keeping records	45
	4. Data analaysis	47
VI	Using evaluations to reassess research and development policies: a retrospective look at rice programs in the Ivory Coast	48
VII	Central Monitoring and Evaluation	63
VIII	A review of ex-post and ex-ante evaluation analysis	105
	1. Some conumdrums	105
	2. Methods and procedures for evaluating agricultural research	114
References		153

List of Tables

		Page
1.	Production of main food crops by value or quantity	6
2.	Value of gross marketed production	7
3.	Estimated research expenditure compared with the value of marketed production	8
4.	Agricultural research expenditure by organizational component	9
5.	Agricultural Research Manpower	10
6.	The logical framework	29
7.	Example of logical framework: career plan	31
8.	Example of logical framework: research program	32
9.	Data gathering methods	41
10.	Production and Consumption of Paddy and Imports of Milled Rice	49
11.	Logframe for Rice Program 1966-70	50
12.	Distribution of New Inputs by SATMACI-RIZ	52
13.	SATMACI Rice Program: Expected versus Actual Achievements	52
14.	Intermediate Costs from Paddy to Rice at Retail Level	55
15.	Principal Rice Statistics 1970-77	56
16.	Gross Production of Principal Food Crops (1000 T)	58
17.	The Plan Paddy	59
18.	Scientific Manyears Allocated to Principal Crop Programs 1983	61

PREFACE

This document was prepared for the Regional Workshop on Research Program Evaluation hosted by the Bangladesh Agricultural Research Council. It has two purposes:

1. To provide the participants with material which will be used during technical sessions to prepare a general scope of work for program evaluations in Bangladesh (chapters I and II) and to select indicators of achievements for a research program (chapter III).

2. To provide the participants with information on evaluation planning and methods for future reference.

Chapters III and IV are written by Dr. Josette Murphy, ISNAR, who also prepared chapters I and V on the basis of material drafted by Dr. Jane Roth, consultant to ISNAR, for an ISNAR internal document. Dr. Howard Elliott, ISNAR, wrote chapter VI.

Chapter VII is reprinted from course note CN-318 from the Economic Development Institute, IBRD (November 1980).

Chapter VIII is reprinted from a report by G. Edward Schuh and Helio Tollini to the CGIAR (Costs and Benefits of Agricultural Research: State of the Art, and Implications for the CGIAR), later published by the World Bank as Cost and Benefits of Agricultural Research: The State of the Arts, Working Paper No. 360, 1979.
Permission to reprint is gratefully acknowledged.

I CHECKLISTS FOR PREPARING A SCOPE OF WORK

This chapter presents checklists of the type of information needed for evaluation of a research program. ISNAR staff use similar checklists to help them decide what to include in a scope of work for a review of a research system or program. The checklists in this chapter were adapted for evaluations of research programs and will be used by the Workshop participants to prepare a general scope of work applicable in Bangladesh.

A program is defined here as a coherent set of complementary research activities which are expected to satisfy a specific objective. The objective can be an identifiable product (a variety, a tool) or more appropriate cropping system or management practices. It cannot be a general goal such as increasing food production, which would be identical for a number of programs and which need actions outside research to be achieved. It should be very clear that the expected product of a research activity must be attainable as a result of the activity.

One cannot evaluate an activity without placing it in context. The scope of work for an evaluation will therefore cover several topical areas, some dealing directly with various aspects of the activity, some dealing with the context in which it takes place, and some dealing with changes the activity is expected to bring about.

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

- The country setting
- Structure, organization and place of the research program
- Planning and budget
- Human resources
- Facilities, equipment and supplies
- Scientific activities and achievements
- Management of the program
- Communication linkages
- 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. What really matters is that the areas needing coverage go beyond the implementation of research itself, because an agricultural research program cannot be assessed in isolation from the agricultural context of the country and from other existing research. An evaluation covers elements from the national situation in which research results will be applied, and specifically includes the policy environment which might influence program implementation or adoption of research results. It should also consider linkages between researchers in the program and other (national or foreign) researchers, national policy makers, and farmers as well as extension and related services.

Other areas covered are those necessary to describe the structure, resources, and implementation of the program. A research program can be implemented within an administrative unit, or it can involve scientists from several divisions in one organization, or from several organizations. It can be implemented in a single location or be multi-locational. It does form an identifiable entity because it has one or more specific objectives.

It cannot be evaluated fairly without knowing the constraints under which it operates (budget, available manpower, rules and regulations) and the agricultural context which its research activities are to serve. An assessment of a program will include analysis of its performance in programming and managing its research activities, and in optimizing the use of its human and financial resources, as well as in the scientific validity of the research itself. The appropriateness of the research in relation to national needs and priorities will also to be considered.

Evaluating an individual institute, a specific program or an entire research system all entail looking at the same nine areas defined above. The difference is one of emphasis: when looking at the research system as a whole, more attention will be paid to national policy and economic factors than when evaluating the implementation of a specific program, for which only those policies or economic factors directly influencing the program would be relevant.

For each of the topical areas, a narrative explaining its relevance for the evaluation is followed by a list of key questions for consideration by an evaluation team. Each list was written to help identify all major issues under the respective topics. It should be very clear that at no time ever would a team be expected to address all these questions. Rather, the team members review the lists during the planning stages and select, add, delete, and modify questions with respect to their particular terms of reference and other relevant factors. The lists are merely catalysts. When deciding whether to include an item on the list, two points should be discussed:

- how would that information be used? It should not included unless it can be justified by being used. "Being nice to know" or "interesting" is not a valid justification

- how much work is involved in obtaining it? Whether it is worth the time and trouble depends on how much the information is necessary to the analysis.

For some questions, a table grouping data for comparative purposes is given as an example of possible format. Team members will find it useful to elaborate some kind of summary presentation of data when appropriate: it facilitates data handling, and the very fact of discussing what to include in a table help the team decide what information is really necessary. There are repetitions and overlap between topical areas, because some questions pertain to several topics and in that case are listed under each one to facilitate selection.

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 for a program evaluation depends in part upon its purpose, particularly whether it will be used in communications with:

1) foreign colleague or donors: they need background data on the country's agricultural sector and administrative organization

2) national policy-makers: they may need precise information on overall research budget and personnel.

Any document prepared for general release needs more information on the country setting than an internal document does. The list given below is more extensive than what is likely to be required and much of it would not be specifically mentioned in an internal evaluation report. It is given in extenso here to illustrate how much knowledge of the country is expected from evaluators. The selection of points from the list to cover explicitly in the scope of work is an exercise in placing the program within the overall agricultural sector of the country. All questions on commodities should be narrowed to the commodity the program deals with.

KEY QUESTIONS FOR INFORMATION ON COUNTRY SETTING (SYSTEMWIDE INFORMATION)

Instructions

Review the following list of questions for relevance to your current evaluation. Not all questions will be needed. Some items should be deleted; some modified; and some new items added. These questions are merely to be used as a starting point. In addition, this list refers to other lists and tables which may be completed as deemed appropriate.

1. Most recent basic descriptive data on size and population

2. Key economic data

* GNP
* major sectors of the economy - - approximate share of each sector in GNP
* total government expenditure or budget
* approximate share of budget for agriculture

3. What are the agricultural statistics?

* percent of population in farming/agriculture
* percent of national income coming from agriculture
* percent of government expenditure supporting agriculture
* principal agricultural commodities - - amount, production area, yield per unit area, and production value
  * crops
  * livestock
  * fisheries
  * tree crops
  * other agricultural commodities

* Table 1: Production of Main Food Crops by Value or Quantity. Land tenure, distribution of farms by size, and major commodities.

* Table 2: Value of Gross Marketed Production of Agricultural Commodities, and if pertinent and possible, estimated value of non-marketed crops

* Table 3: Estimated Research Expenditure Compared with the Value of Marketed Production for Agricultural Commodities
* principal exports/imports of commodities of plant or animal origin

4. What are the produce prices? (may be limited to commodities potentially affected by agricultural research in the program being evaluated)

* support prices, on what commodities
* relation of support prices to world market prices, on major commodities
* farm gate prices in relation to prices in the market place and in relation to world prices

5. What is the infrastructure?

* credit system
* fertilizers - - availability and prices
* seeds - - availability and prices
* insecticides/herbicides/fungicides/pesticides - - availability and prices
* marketing and storage
* transport systems
* irrigation systems
* flood control
* electrical and other power systems

6. What is the natural resource base?

* map(s) of the country's agro-climatic-topographic-ecological regions with brief descriptions of both current and potential land use
* map(s) of the country showing locations of research and experiment stations and testing sites, and production locations (areas) for major agricultural commodities
* map(s) of country showing population density and major transportation systems

7. What is the long term agricultural potential of the country?

8. What are the current agricultural limitations of the country?

9. Basic organization of the national agricultural research system

* Administrative Chart of National Agricultural Research System
* Linkages Between Research Entities Active in the Country
* What are the current national priorities and how are they determined?
* Is there a planning mechanism at the national level for deriving national agricultural research plan? Describe.
* What are the principal features of the national plan? Is it a multi-year plan?
* Table 4: Agricultural Research Expenditures by Organizational Component

10. What is the national manpower situation for agricultural research?

* What are the estimated current numbers of agricultural research scientists (by discipline, if known) and technicians? Are there shortages/oversupplies in any areas? Where? Why?
* Is there a national multi-year plan for manpower and training? What is it?
* Is the output of the country's education system sufficient for the projected needs?
* Does the country try to recruit scientists and technicians from other countries? How extensively? With what success?
* Do university courses fit the graduates' need for subsequent work in agricultural research?
* Table 5: Agricultural Research Manpower, by Highest Degree, Citizenship, and Research Areas

Table 1. Production of Main Food Crops by Value or Quantity

Commodities, listed by categories	Years				
	19--	19--	19--	19--	19--
Category A					
Commodity A_1					
Commodity A_2					
Category B					
Category C					
TOTAL					

Table 2 Value of Gross Marketed Production of Agricultural Commodities, 19-- through 19--

	Years				
Commodities, listed by categories	19--	19--	19--	19--	19--
Category A					
Commodity A_1					
Commodity A_2					
Category B					
Category C					
TOTAL					

Table 3 Estimated[1] Research Expenditure Compared with the Value of Marketed Production for Agricultural Commodities

	A	B	C	D	E		
		Research Expenditure					
Commodities	Production Value[2]	Government	Private Sector	Contributed International	Total	Percentage of Total Value (Column A)	Percentage of Total Expenditure (Column E)
Commodity A							
Commodity B							
TOTAL						100%	100%

[1]Estimation may be needed to attribute some research expenditures to particular commodities.

[2]Figures appear in Table 1.

Table 4 Agricultural Research Expenditures By Organizational Component[1], 19-- through 19--

Components	Expenditure Categories[2]	Years 19--	19--	19--	19--	Total	Percentage
Component A	Category 1						
	Category 2						
	TOTAL						
Component B	Category 1						
	Category 2						
	TOTAL						
TOTAL	Category 1						100 %
	Category 2						
	TOTAL						

[1] Components may be agencies, institutes, or programs depending on the country's budget process.

[2] Examples of categories are recurrent or development, staff, equipment, commodity group.

Table 5 Agricultural Research Manpower, By Highest Degree, and Research Areas

19-- through 19--

*Categories	National Scientists						Expatriate scientists	Technical Support Staff				
	Year	BSc	MSc	PhD	Other	Total		Total Scientists	University Graduate	Technical diploma	other	Total Support staff
	19--											
	19--											
	19--											
	19--											
	19--											
	19--											
	19--											
	19--											
	19--											
	19--											
Total	19--											
	19--											
	19--											
	19--											
	19--											

* These categories may be research areas, commodities, programs, departments, ecological zones, or some other groupings.

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.

KEY QUESTIONS FOR INFORMATION ON STRUCTURE, ORGANIZATION AND PLACE OF THE RESEARCH PROGRAM

Instructions

Review the following list of questions for relevance to your current evaluation. Not all questions will be needed. Some items should be deleted, some should be modified, and some new items added. These questions are merely to be used as a starting point.

1. What is (are) the objective(s) of the program? Areas of responsibility?

2. Describe the organizational structure of the program. What are the major strengths and weaknesses of the organizational structure?

3. List the institute(s) and their department(s) involved in the program. If several institutes are involved:
   What is the division of authority among institutes? What is the division of labor?

4. To what extent does the structure of the program facilitate or hinder its key functions? research? planning? management? communication? impact? What evidence is there to support this inference?

5. How do the individual components of the program fit together to cover the areas of responsibility or objectives?

6. How are decisions made? centralized or decentralized with regard to organizational operations and structure, and personnel decisions?

7. Where is the program located in the national agricultural research system?

8. How does the program fit into the system, including reporting lines, budget and program approval authority, and review authority?

9. How is the program related to other agencies in ways other than reporting, approval and review, such as: formal and informal interrelations with other government entities or programs and the private sector, including universities, agro-industrial research offices, and independent institutions?

10. Is there any overlap of research responsibility or activity with other research programs? Are there parallel, interdependent, competing, or redundant responsibilities and activities?

11. To what extent is the program part of a vertical dependency relationship? (The most common example is perhaps the dependency of the research program on universities in the host country for providing trained manpower in agricultural research disciplines. Other examples would be where a different agency is responsible for disseminating the research findings of the institute being evaluated, or where a different agency controls all imports, including equipment required to carry out the institutional or program research efforts.) Beyond reporting whether such relationships exist, indicate the extent to which the relationships are working smoothly.

12. How does the program compare with the rest of the research system in budget, staff, other resources, and products and services?

C. PLANNING AND BUDGET

Information needs to be collected on the planning process for the program, as it interfaces with the national planning process and an resource allocation. Program managers and planners will recognize many questions from long and short term program planning and formulation; budget officers will see familiar questions about the structure and timing of the budget process; and evaluators will notice issues of needs assessment.

KEY QUESTIONS FOR INFORMATION ON AGRICULTURAL RESEARCH PLANNING AND BUDGET.

Instructions

Review the following list of questions for relevance to your current evaluation. Not all questions will be needed. Some items should be deleted; some should be modified; and some new items added. These questions are merely to be used as a starting point.

1. What are the national priorities for agriculture and for research and how are they determined?

2. What are the program objectives? How were they determined? How do they relate to national priorities? Are they realistic? Are the needs and priorities of all relevant parties (government policy and planning offices, intended beneficiaries of the institute or program, etc.) well articulated?

3. Assuming realistic objectives have been set in general terms, what process is used for deciding what work to do and who makes the decision?

4. Is there a mechanism for program formulation which includes user representatives? Does the existing research program match the identified needs?

5. Is there an effective planning mechanism for the allocation of resources to different stages of the research process? How are funds allocated between research in different fields of science and technology?

6. What is the budget of this program? How was it formulated and presented? Is there a PWB (program within budget) system?

7. Has the availability of support from external sources affected the design and implementation of the program?

8. After approval of the budget, how were funds released for capital and recurrent expenditures? How are yearly budget and workplan prepared? What is the timetable for preparation? presentation? approval? release of funds?

9. What tendering procedures must be followed, especially for capital works and major items of equipment?

10. Are funds released early enough for the field programs? Is there any delegation of authority for disbursing funds to the point of action?

11. Is there continuity of funding for recurrent expenditures other than staff salaries?

12. To what extent is there multi-year planning? How extensively are unit managers or supervisors involved? To what extent is the multi-year plan followed? What aspects of the program are specified in the multi-year plan?

13. Is there a true costing out procedure in the budget process or is it really just a level of effort activity?

14. What are the institute's or program's procedures for budget expenditures?

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. The data are organized to the extent possible in a form that allows comparisons with data on the supply of such manpower graduating from host country universities and the manpower data for the agricultural research system as a whole.

KEY QUESTIONS FOR INFORMATION ON MANPOWER AND TRAINING

Instructions

Review the following list of questions for relevance to your current evaluation. These questions may also apply to the university complex. Not all questions will be needed. Some items should be deleted; some should be modified; and some new items added. These questions are merely to be used as a starting point.

Manpower Situation

1. See section on national manpower for agricultural research in the country setting checklist.

2. What are the current numbers of research scientists, field and laboratory technicians, and administrative staff in the program? By levels of training, areas of specialization, and nationality (national and expatriate)? Where are there current shortages or oversupplies? What are the projected needs by ...?

3. What is the turnover history?

4. What are the current numbers of research scientists, field and laboratory technicians, and administrative workers being trained (in universities) for positions in agricultural research? Is the supply adequate for current needs? projected needs?

5. Are there arrangements for contract work, and international and regional cooperation?

6. Are the total resources of manpower adequate for a realistic program of research to be carried out on a continuing basis? For fulfilling the program's objectives?

7. What are the salary levels of research scientists, field and laboratory technicians, and administrative staff in the program? Compared with comparable positions in the private sector?

8. Is the remuneration of all grades, including supporting service staff, both absolute and relative, to administrative and executive staff, adequate to attract and retain staff in sufficient numbers and of good quality?

9. How many research staff in active research positions have the five or more years experience?

10. Is there appropriate staff assessment? Has it been effective? Related to recruitment and training? How? Cite examples.

11. Is there any effective merit promotion system to attract and retain staff?

12. Is sufficient use made of well-trained administrative support staff to manage those parts of the system which do not require specialized scientific knowledge, thus allowing the specialists to concentrate on their research work?

14. Is there active interaction at the post-graduate level between the educational system and the program? Effective? Evidence?

15. How are scientits and techniciens selected for the program?

In-service Training

16. What in-service training is available? Describe the nature of training and the number and type of staff (i.e., professionals, technicians) involved. How good is it?

17. What training is provided for managers and supervisors? Describe the nature of training and the number of staff involved. Is the training adequate? Effective?

18. Is there a multi-year plan for manpower and training?

E. FACILITIES, EQUIPMENT AND SUPPLIES

Information on the number, size, and condition of facilities, equipment, and supplies 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.

KEY QUESTIONS FOR INFORMATION ON FACILITIES, EQUIPMENT, AND SUPPLIES

Instructions

Please review the following list of questions for relevance to your current evaluation. Not all questions will be needed. Some items should be deleted, some should be modified, and some new items added. These questions are merely to be used as a starting point.

Status of Facilities, Equipment, and Supplies

1. What is the location of the facilities and research activities? Proximity to production areas? Obtain map(s) of the country showing the location of research and experiment stations and testing sites of the program, and the locations (areas) for major agricultural commodities. See Country Setting questions for related national maps.

2. Is the program(s) established in appropriate geographical locations?

3. Are the total resources of facilities, equipment, supplies, and operating budget adequate for a realistic program of research to be carried out on a continuing basis? If not, what are the most crucial needs?

4. Are the stations adequately provided with land, buildings, and equipment to insure regular residential staffing and full operational efficiency?

5. What laboratory facilities are available? how are they equipped?

6. Is the equipment in good working condition? Is it reasonably current?

7. What library materials are available? Are they adequate for all areas of research being conducted? Is IARC and international information included? What additional materials would be most useful? Are materials located in a central library or distributed? Describe. How accessible are library materials? To substations?

Use of Facilities, Equipment, and Supplies

8. Are staff adequately trained in proper use and maintenance of equipment and supplies?

9. To what extent is equipment left idle? Are there instances of oversupply of some types of facilities, equipment, or supplies?

10. Is any research not being done due to lack of proper equipment and facilities?

11. Are any facilities or equipment currently being shared? Have the opportunities for sharing facilities or equipment with other organizations ever seriously been investigated?

13. How much are the library materials used? By whom? To what extent are appropriate staff aware of the materials?

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.

KEY QUESTIONS FOR INFORMATION ON SCIENTIFIC ACTIVITIES AND ACHIEVEMENTS

Instructions

Please review the following list of questions for relevance to your current evaluation. Not all questions will be needed. Some items should be deleted, some should be modified, and some new items added. These questions are merely to be used as a starting point.

1. Were the program's activities implemented as planned? Describe. Cite examples.

2. To what extent have the agricultural research activities of the organization kept pace with the planned schedule? Are revisions of the time line now needed?

3. Have new problems been encountered calling for serious revisions in the research plan?

4. To what extent are the field and analytical research methodologies effective?

5. It is often extremely useful to select a few protocols or very specific projects and retrace their history, from conception to utilization of results within the program as a whole. This is one exercise where even a brief evaluation should go into details, because they will illustrate many aspects of implementation:
   - what is the rationale behind this protocol? how does it relate to program objectives?
   - who designed it?
   - who in reality set up the experiment?
   - who observed the results? how?
   - who reported the results? how? to whom?
   - who interpreted those results?.

   It is not unusual to discover that the researcher did not fully control the experiment nor observed the results. The validity of the experiment, quality of observations, and accuracy in reporting should then be scrutinized in a few actual cases.

6. What are the principal accomplishments (technological innovations) in recent years? What products, processes, and new information result from research? Describe, give all characteristics and compare with indicators of expected results.

7. To what extent have the accomplishments of the program been recognized outside the organization? In the agricultural research community? In industry? In the government? by the general public?

G. MANAGEMENT

In a broad sense, management encompasses almost all areas of a program's operations. It certainly includes the area of planning and budget and bears directly on resource availability and utilization, and on inter and intra organizational communication, areas taken up in the following sections. In this section, a special set of management issues are addressed -- those concerned with decision making, quality control and resource management as they relate to meeting stated objectives. 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. This list should be considered an addition to each of the other lists.

KEY QUESTIONS FOR INFORMATION ON MANAGEMENT

Instructions

Review the following list of questions for relevance to your current evaluation. Not all questions will be needed. Some items should be deleted; some should be modified; and some new items added. These questions are merely to be used as a starting point.

Decision-Making Structure

1. Where are resource level and allocation decisions made for the program? Who decides? Who influences? Describe the nature, course and timing of the most recent major decisions of this sort.

2. Where are the programmatic decisions made for the program? Who decides? Who influences? Describe the nature, course, and timing of the most recent major decision of this sort. Do this for the program as a whole and for each of its major components.

Resource Management

3. Is there an effective planning mechanism for the allocation of resources to different stages of the research process? How are funds allocated between research in different areas? How are resources allocated and used within the program?

   * research areas
   * manpower
   * facilities, equipment, and supplies

4. To what extent does the program contract for research (as distinguished from in-house research)? International and regional cooperation? How effective is this work? Cite examples.

5. Is the remuneration at all grades, including that of supporting service staff, adequate to attract and retain staff in sufficient numbers and of good quality? Describe.

6. Is the best use made of staff energy and skills? Is sufficient use made of well-trained administrative support staff to manage those parts of the research which do not require specialized scientific knowledge, thus allowing the specialists to concentrate on their research work?

Quality Control

7. Who has primary responsibility for the quality of research in the program? How is the responsibility exercised? To what extent does management actually monitor program progress and expenditure? Is there a formal review or evaluation procedure? Is there a staff assessment system? Is there a quality control procedure? Is peer review used?

8. Who conducts programmatic and budget reviews and evaluations? When done most recently? How active? To what extent is the information used? By whom? Cite examples of use.

9. At the research unit level, is there a mechanism for maintaining objectivity in work and for preventing programs from becoming too diffuse? Periodic assessments? If so, what kind, by whom? Cite examples.

H. COMMUNICATION LINKAGES

Agricultural development and increased productivity 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 linkages with other research organization within and outside the country, development agencies, and policy makers.

Unfortunately, these communication linkages are often difficult to assess in a reliable fashion because informal relations often exist where offical ones do not. For this reason, key questions in this area make specific distinctions and request examples ("active communication", "effective communication", "can you give some examples").

KEY QUESTIONS FOR INFORMATION ON COMMUNICATION LINKAGES

Instructions

Review the following list of questions for relevance to your current evaluation. Not all questions will be needed. Some items should be deleted, some should be modified, and some new items added. These questions are merely to be used as a starting point.

1. Describe the communications across disciplines within the program. Are these communication linkages active? effective? Cite examples.

2. Describe the linkages with: other related institutes or programs; international organizations; universities, research activities outside the public sector. Is the communication active? effective? Are the communications occurring at both the scientific and managerial levels? What means of communication are being used?

3. Describe the linkages (formal and informal) between research units, field extension service and farmers. Are they active? Effective? Cite examples. How well informed are researchers concerning the conditions (problems, technology, methods and level of knowledge) at the farm site?

4. Are there links with policy-making levels in the controlling Ministry? What means of communication are used? Cite examples.

5. Are any efforts made to inform planners, politicians, or the general public about the progress and potential benefits of the program's research? How? How effective? Cite examples.

6. How are research results disseminated? to other researchers? to extension agencies? to farmers?

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.

KEY QUESTIONS FOR INFORMATION ON CONTRIBUTION OF THE RESEARCH PROGRAM TO DEVELOPMENT

Instructions

Review the following list of questions for relevance to your current evaluation. Not all questions will be needed. Some items should be deleted; some should be modified; and some new items added. These questions are merely to be used as a starting point.

1. How have research accomplishments of the program been disseminated? to whom?

2. For each accomplishment, can it be expected at this time to show tangible effects on production? On the population?

3. If so, what has been the effect of these accomplishments at the farm site? For the country? How extensively have innovations been adopted? By which categories of producers? Can this effect be expected to grow? What factors, if any, have constrained adoption? What is needed to overcome these constraints? Are there social or political implications? Be sure to look into a very broad range of possible factors: agro-ecological, social (household structure, labor force), economic (marketing channels, relative prices of commodities, availability and cost of inputs), policies such as input/export regulations or organization of extension and banking services, infrastructure. See the chapter on Using Evaluations for example of interactions between different aspects of the economy and their influence on adoption of research results.

5. Do some of the activities of a program seem to be having a greater impact than others? Describe.

6. Who are the principal beneficiaries of the results? how do they benefit? Are they the intended beneficiaries envisioned in the institute or program mandate?

7. How do these contributions compare with the expected causative line from output to purpose and to goals in the program design?

II EXAMPLE OF AN EVALUATION SCOPE OF WORK

This chapter presents exerpts from the scope of work for evaluations of research programs in Indonesia. It was developed with the help of the ISNAR checklists presented in chapter 1.

1. The team will review the program activities and management of the program of AARD

2. The primary purposes of the review are: (a) to provide the Government of Indonesia, AARD in particular the relevant institute directors with an analysis of the past, ongoing and proposed activities of the research program; (b) to identify ways and means of strengthening the research program; and (c) to increase the in-house evaluation capacity within AARD.

3. The review will form part of a series of about ten reviews which will eventually cover all of AARD's activities and which will examine both the achievements of the research programs to date and their objectives for the period until 1990.

4. The review is expected to report on the past, existing and proposed programs of the research program and to make recommendations with respect to:

   (a) their management
   (b) the quality and relevance of the current and proposed research
   (c) the adequacy of the human, physical and financial resources
   (d) the effectiveness of the linkages of the Institute with the scientific establishment both in Indonesia and overseas
   (e) the nature and effectiveness of the linkages with the extension services and other agencies providing services to agriculture
   (f) possible new areas for national, regional and international support

5. Each review team will be expected, so far as is practical and relevant, to report within the framework of the attached outline so that its report can be incorporated into a global overview at the end of the series of reviews. For each program reviewed additional "specific" terms of reference will be drawn up.

6. A specific scope of work will be drawn up by the Director responsible for each crop program.

Each scope is expected to follow the outline below:

1. BACKGROUND (This section to be drafted by the review team secretariate and to be available for the team's initial briefing)

Agriculture in the economy of Indonesia

Agricultural goals of the latest 5 year plan

The socio-economic importance of the crop in the Indonesian economy

The status and role of AARD

The role of the research institute(s) concerned with the crop

The physical resources available to the program:
- location of stations and farms
- status of lands and equipment

The human resources available to the program:
- existing numbers and level of training
- number currently in training, wastage
- training targets for 1990

The financial resources available to the program:
- current budget
- domestic/foreign components
- past and future trends

The specific objectives of the program research:
- historic
- current

2.A PLANNING AND PROGRAM FORMULATION

PRIORITY SETTING:

- What are the priorities for research and the mechanism for determining them? What kind of consultation (with policy makers, private sector, extension services and farmers) is involved in this process?

- What criteria are used in setting the priorities? To what extent do these take account of available resources (human, physical and financial)? Do they take into account information based on economic analysis or empirical models?

PROGRAM FORMULATION:

- How and who decides what research activities (projects) will be carried out in each priority area.

- How does the program relate to the resources available to the institute? Is there an appropriate "program of work and budget" linkage?

- How are the individual units within the institute coordinated in multi-disciplinary research projects.

RELEVANCE OF THE PROGRAM IN TERMS OF THE OBJECTIVES:	· How does te current program relate to the objectives and priorities? Are there any major objectives not covered by the current research program?

2.B IMPLEMENTATION

PROGRAM ACTIVITIES:	· What research is being done? What is the main commodity, ecological or disciplinary emphasis? Is the research relevant to the goals of agricultural development? Is there duplication elsewhere in AARD? How does the physical location of the current research program relate to both agricultural development requirements and resources available for research?
METHODOLOGY:	Is the methodology being used appropriate in relation to the research objectives and to the resources available?
RESOURCE AVAILABILITY:	Are the total resources (physical, financial and human) adequate for a realistic research program to be carried out on a continuing basis? Is the manpower pool available appropriate for the research program in terms of experience level, disciplinary expertise and support staff mix? · Are funds available on schedule? · To what extent is present program execution dependent upon external financing.
RESOURCE UTILIZATION:	· To what extent does the research program effectively utilize the funds, lands and manpower available to it each year. · Are the facilities and equipment being well maintained? · To what extent does the Program sub-contract research and with what results.
MONITORING AND EVALUATION:	· Are there any formal monitoring and evaluation mechanisms being used by the Program with respect to progress, expenditure control, and staff assessment?

2.C LINKAGES

WITHIN AARD:

- What channels of communication exist -
  (a) within the program
  (b) with scientists in other AARD programs
  - and how effective are they in both directions?

WITH OTHER RESEARCH ORGANIZATION:

- What channels of communication exist with -
  (a) other Indonesian research institutes
  (b) universities
  (c) private sector
  (d) international institutes,
  - and how effective are they in both directions?

WITH EXTENSION SERVICES AND FARMERS:

- What channels of communication exist with -
  (a) field extension workers
  (b) farmers
  - and how effective are they in both directions?

WITH POLICY MAKERS:

- What channels of communication exist with -
  (a) AARD headquarters
  (b) AAETE
  (c) DG's
  - and how effective are they in both directions?

2.D IMPACT

RESULTS TO DATE:

- How much of the program undertaken during the past 3 years has yielded tangible results?
- What are these results?
- Do the results meet farmers needs or are they likely to when the research is more advanced?

USE OF RESULTS:

- To what degree have the results been adopted by farmers.
- What are the main constraints to adoption?
- Is research being undertaken to overcome these constraints?

3. RECOMMENDATIONS

A. Program planning

B. Research activities

C. Resource utilization

D. Linkages

E. Cooperative activities

III. INCORPORATING INDICATORS OF ACHIEVEMENTS INTO PROGRAM DESIGN.

The importance of systematically planning for evaluations when designing a research program is emphasized throughout the workshop, because actual achievements of a program can only be evaluated by comparing them with "standard" achievements which the program was designed to provide. This requires that indicators of achievements be identified and clearly stated when the program is being designed. The very fact of having to specify standard achievements early on is an excellent discipline. It forces the program designers to clearly define program objectives, identify expected results, and specify which indicators will be used to measure actual results and compare them with what was expected.

This section introduces the steps related to evaluation needs which should be taken when designing a program. It suggests a format (see example below) which most readers will recognize right away as being very similar to the Logical Framework used by the United States Agency for International Development 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 uses implicitly whenever making a plan of action, whether for a research program or for a personal decision.

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 achievement of each level. The four rows will be described 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".

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 changes for success.

TABLE 6. THE LOGICAL FRAMEWORK

	Narrative Summary	Verifiable indicators	Means of verification	Important assumptions
Then →	GOAL			
If Then →	PURPOSE			
If Then →	OUTPUT			
If	INPUT			

To emphasize the universality of the thought process an example from personal life is now given. Suppose a student has just completed a Bachelor of Sciences degree and wishes for a career in agricultural research, with the ultimate goal of becoming director of a research institute. The student's plan may look like this (Table 7).

The student has only a BSc so he or she must first develop his knowledge and competence (verified by the award of a graduate degree), and also develop his management abilities, so that he can achieve his purpose of being able to obtain a research position with leadership potential. The ultimate goal is a position as director of an institute.

An evaluation of the student's progress towards his/her ultimate goal after three years may find that he no longer is a student at the university yet did not receive a PhD degree. How to interpret this lack of achievement depends on its cause. If the student did not do the required work and flunked, the fault is his (insufficient input). If he had to withdraw in spite of its good standing because the program of financial assistance he depended on was curtailed, the assumption about financial assistance was incorrect. In this second case, the evaluation results carry an important warning for the government: insufficient funds are causing a loss of human resources for the country.

An evaluation of a research program would use primarily the input, output and purpose narratives, and the two central columns (objectively verifiable indicators, and means of verification) (Table 8). 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

TABLE 7: EXAMPLE OF LOGICAL FRAMEWORK: CAREER PLAN

Narrative Summary	Verifiable indicators	Means of Verification	Important assumptions
GOAL career as agricultural researcher, all the way to institute director	regular professional growth	job promotions publications invitations to conferences and committee work	career patterns on competence stability in staff in institutes
PURPOSE become able to get job with career potential in research institute	suitable job within 6 months of degree	job contract title and job description	demand for researchers continue
OUTPUT knowledge and research competence + management abilities	PhD degree successfull involvement in university's student committee	- university transcripts - degree with 5 years - dissertation published - elected secretary of student cttee	- find program well adopted to needs -
INPUT work time for studies money time for complementary activities	passing grades MSc within 2 years	- university transcripts - MSc degree - reports by professors	- get accepted in graduate school - obtain financial assistance

TABLE 8: EXAMPLE OF LOGICAL FRAMEWORK: RESEARCH PROGRAM

	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

program, and analyzed at regular intervals to satisfy reporting requirements. An evaluation at the end of the program will then have most of the necessary material to measure the indicators of achievements. This is not sufficient however, because a good evaluation should interpret its findings, analyzing 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 betwen 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 that 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 consistently yields more than 2 tons/ha in real farm conditions. Means of verifications 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.

IV. SUGGESTIONS FOR PLANNING AN EVALUATION

This chapter presents some general principles and guidelines on planning an evaluation. They are presented around six key requirements: clear evaluation objectives, definite but flexible plans, efficient use of existing information, purposeful field visits, quality controls, and teamwork.

1. Clear Evaluation Objectives

When planning an evaluation, it is essential to first clearly establish who will use the results and what they will use the result for. This is essential because it determines a number of factors which influence the evaluation activities:

- It determines appropriate recommendation domains and therefore necessary data: policy makers don't want details on technical aspects of research but may need information on the production potentials of proposed research programs and relative economic and social costs. On the other hand, the director of an institute needs to know this technical information if it poses new requirements for staff or laboratory facilities.

- Who uses the results and for what purpose also determine the level of accuracy needed in data gathering. Making the extra effort to obtain precise data is a waste of time and resources if the data are used for only rough approximations.

- Finally, the way the data are to be used also determines the time frame within which the evaluative activity must take place. Nothing is more useless than an evaluation report published six months after a decision had to be made.

The same evaluation can have several categories of users, but it should be clear which category is the primary audience.

2. Definite but flexible plans

Fairly precise plans should be made ahead of time on data to obtain, people to interview, and sites to visit, but sufficient flexibility should remain possible to allow following up a new line of inquiry suggested by a first analysis of preliminary data. It is important not to limit artificially the categories of data relevant to the job. For example, studies of research systems in recent years have emphasized the importance of government policies related to prices and exportation for adoption of research results. It is also essential to check the information gathered and the tentative conclusions derived from them against each other and not to accept any one point of view at face value. This goes for already existing data (statistics, reports) and for information gathered as part of the evaluation.

However, focusing on getting more data until the last minute is a common problem in all evaluative activities. It is always tempting to get just one more piece of information, another document, or add one more question

to a survey. This helps to reassure the evaluator that he/she has a solid base of data. Yet it is essential to allocate enough time for analysis of the data and for the "absorption" process allowing the team to sift through possible relations and potential recommendations. This has to be an iterative process, with preliminary analyses.

There can be no standard list of data to be obtained, since the boundaries and characteristics vary with each research system. The suggested data checklist listed in Chapter I are indicative of what may be needed, but specific lists will have to be made for each evaluation, given its purpose and circumstances. These topical areas are not mutually exclusive but they have been found convenient to group the various information which could be necessary, as a first step in preparing a scope of work for an evaluation.

The purpose of a scope of work is to indicate what is to be done, by whom, and what is the expected outcome of the work. In some situations it may be important to indicate the limits of the study. In general, the scope should be sufficiently broad and flexible to avoid constraints to evaluation activities but sufficiently detailed to inform the evaluation team of the main items to be given detailed study.

As with all data collection efforts*, items included in the scope of work should be:

- necessary: not just interesting to know, but really required for the desired analysis. Time is too precious to spend gathering information which will not be used.
- neutral: written in such a way that it does not influence the answer.
- clear and specific enough to be understood in the same way by everyone and to indicate the specialities of the staff needed to do the work.
- conducive to the desired analysis with as little additional handling as possible.

When preparing a scope of work, it is often helpful to think about the likely table of contents of the report in order to identify the categories of information to be included. At this stage, however, the order in which the sections will be presented in the report is not important. This will derive from the findings which determine which issues need emphasis. The order in which the items to be covered are listed is not, in any case, the order in which the information is obtained. This is a matter of convenience.

When very little is known ahead of time, the scope should remain very broad and flexible so as not to casually reject what could turn out to be a useful line of inquiry. It should be comprehensive enough so that the tasks of each team member can be specified, without being too limiting, because a new area of concern could emerge during fieldwork, and the team should have the flexibility to pursue it.

---

* Suggestions for preparing questions for data collection are given in Chapter V.

3. Efficient use of existing information

A visit by evaluators is always disruptive, but it is particularly annoying to researchers and program leaders to be asked to describe the program. Questions about clear and absolute facts should be answered from available documents, it is a waste of time for all concerned to use interview time or field visits for this. However, not all facts are "clear and absolute," data about budget disbursement, staff involved, experiments in progress cannot simply be taken from program documents since they may differ from what was planned. Basic documents on the economic and agricultural situation of the country should be gathered at this early stage, as well as reports and descriptive documents on the research institutes' legal status, mandates, structures and activities and on current research and development objectives and priorities. Statistics on production, official prices of commodities, boundaries of agro-ecological zones can be taken from published data, as well as past dates in the history of the program.

It is important that the evaluation team members have as much of this type of information made available to them before the evaluation itself (i.e., interviews and field visits) begins. This work of compiling data is best done by a team assistant, someone who is given access to all necessary sources of information (library, archives, statistics and publications from various agencies) and understands what information is required. The work of the assistant may involve compiling tables from data from various sources. The assistant should work under the authority of the evaluation team leader. Once all this material has been compiled, it can be reviewed by all team members, who will be in a better position to finalize the evaluation's scope of work, decide which field visits to undertake, and whom to interview.

4. Purposeful field visits

Scheduling field visits and interviews can be difficult, especially for complex programs involving several institutes and field stations in several locations. It is not appropriate for the host institute(s) to decide on the evaluation team's itinerary and schedule. This is partly because they very naturally would rather show their best case than the worse one, and also because they are likely to conceive of the visits as "seeing buildings and fields", when what matters for the team is to talk to the people who implement the program in conditions where they feel free to explain how the station really functions from day to day.

Here again, the optimal evaluation schedule depends on the particular circumstances, but in general a smaller group gets a more in-depth understanding of how a station actually operates than a large group of visitors. It does not necessarily mean that the full team must follow different itineraries, but at least that each individual visit should be for a small group of visitors. The more usual scenario where the station director makes a presentation while his researchers stand respectfully aside is a useful beginning but is not sufficient. It is essential that team members speak with the researchers informally. This can be done quite naturally during visits to the labs or field, when one team member stays with the "official" party and the other talks to the staff.

The team needs to hear more than the official speeches, and it needs time every day to discuss and reflect on what is being learned. The schedule should not be too full. Time should be left open for more informal discussions with researchers and technical assistants on a station or in a laboratory, as well as for discussion among the team and writing up notes for the day.

Furthermore, research activities need to be analyzed in context, so the team may need to meet with local officials of the extension service, development agencies, and other entities which are instrumental in identifying research problems and adoption of improved technology. It can also be enlightening to speak with a few farmers, both relatively well off (likely to be identified as "progressive" by the extension service) and some with minimal resources. It is not enough to tell the host institute that team members want to meet with farmers, because it may simply summon the village's richer farmers to the station to be interviewed in due course, most likely with research or extension staff present. In that case, all the team members are likely to hear is the farmer's understanding of the official extension policy. From a sampling point of view, stopping by in between station visits to speak to people working in the fields, in their houses or at the tea house is more likely to give a correct understanding of how much farmers in the area are aware of research results, what their main problems are, and what are constraints to improved practices.

Getting a group of farmers, including some older ones, to discuss among themselves what has changed in farming practices and cropping systems over the last 10-15 years, why, and what would need to change to achieve significant increase in production is most enlightening. It is not likely to happen, simply because of lack of time and flexibility in schedule, unless the team leader insists on it before an official program is agreed upon.

5. Team preparation

Because of the scope and nature of evaluative activities in agricultural research, several individuals from complementary backgrounds are involved. However, a large team (6-7 people) is more difficult to manage, both logistically and intellectually, and is not likely to be more efficient. A team of four or five seems optimum if many field visits are expected: it allows for a split into two subgroups and yet does not make overall discussions too difficult. The nature and amount of work to be done will determine how many persons are needed, their specializations, and will influence the selection of the individual members of the team. In any case, all team members must participate in the total duration of the main fieldwork period. Having someone come for only part of the fieldwork disrupts the process of iterative analysis of options which is indispensable for appropriate recommendations.

Many factors need to be considered when selecting team members: disciplinary and commodity expertise, experience with planning and management, current affiliation, past position in the country, contacts with national officials, and ability to work in an interdisciplinary effort. While involvement in the program being evaluated is not necessarily a problem, it is essential to have several "outsiders" in the team (whether national or foreign), who will bring fresh questions and views to the evaluation.

The role of each member in the team should be clearly specified, the division of labor during fieldwork should be clarified, as well as what exactly is expected from each member: only notes and conclusions, draft of a section of the report, or final text of an annex?

Even with an experienced team, a fairly structured set of meetings with the entire team before fieldwork begins is necessary. It should include:

- discussing the background information available, key issues identified, and preparation of the detailed scope of work. This is already an analytical discussion, during which relevant experience from other programs and countries can be mentioned
- on the basis of data available at that stage and of the scope of work, a list of information still required may now be drawn
- discussion of division of responsibilities within team and of logistics for site visits (how the team will divide)
- discussion of the (tentative) format for the report.

It has been found helpful to begin fieldwork with a formal briefing of the team by the leader and senior researchers of the program. This serves several purposes. For the researchers, it is an opportunity to behave as partners to the team rather than as an object of evaluation. For the team, observing the interaction among the research staff (who speaks, who sits where) and which issues they choose to emphasize is as enlightening as the content of the briefing itself. The briefing should not be limited to a description of the program structure and resources but present current research activities and results to date, as well as eventually the researchers' views on problems and constraints to achieving objectives. What is left unmentioned, even though the team may be aware of it, can be as important as what is presented.

However, further meetings with the program staff as well as meetings with relevant other organizations and individuals should be held in small groups, with the team split by groups of two. These meetings should be as informal as possible with lower rank officials, so that they feel free to say what they think and how their program functions on a day-to-day basis. To reach that level of confidence may require that the team meet with some individuals in the absence of their superiors, something which may be possible only outside regular office hours. These and other approaches to interviews and meetings are presented in more detail in Chapter V, together with suggestions on how to insure that the information gathered is circulated within the team and discussed regularly.

6. Quality Controls

Data collection plans should have some built-in checks for accuracy and completeness. The concept of triangulation is key throughout an evaluation: the operations of the research program are viewed from several perspectives. Related information is gathered from more than one source and reviewed by more than one team member. Whenever possible at least two individuals in comparable positions are interviewed concerning

the same issue and people in different positions are queried about the same issues. Also, interview information is corroborated by observation; and documents are checked for validity and currency. The justification for such extensive triangulation is that the team needs to be as certain as possible that the information obtained is valid and reliable.

In this same pursuit of quality information, team members work together to develop and standardize data collection procedures as much as possible. The idea is to obtain consistent and accurate information pertaining to the mission objectives. There simply is not time for collecting much extraneous information. The quality of data is further enhanced by writing clear, concise questions that will be fully understood, by planning the interview procedures to improve the communication exchange, and by developing interview and observation guides to improve the integration of the evidence gathered by team members. Team briefings and periodic team meetings during the evaluation also improve effectiveness and provide a forum for discussion and critique of findings.

V GUIDELINES FOR NON-EXPERIMENTAL DATA COLLECTION

This chapter presents some methodological guidelines on the use of non-experimental observations and of interviews during an evaluation. They are adapted from the general evaluation methodology and are applicable to any type of non-experimental data collection. Observations and interviews are discussed separately for ease of presentation, but in real fieldwork they constantly intermingle: observing the person interviewed is as important as listening to his/her answer. The strengths and weaknesses of these and other means of obtaining information are summarized in table .

This chapter also includes some suggestions for keeping records from observations and interviews, and for analyzing these qualitative data.

1. Observations

Observation is one of the major investigative tools in evaluations. Team members observe first hand the conditions in the institute(s) and stations visited. Team members can observe agricultural research programs in operation, research facilities and experiment sites, equipment in use, and possible extension/outreach activities. Team members can view interactions among people, note physical surrounding, and see a full range of farm and station activities. Observations can also be used to verify earlier inferences and uncover new issues.

Observations are particularly useful as process checks. Is the program as implemented faithful to the proposal description? Does the program have the necessary means (e.g., personnel, equipment) to meet its objectives? What is an accurate description of the program in operation? Since observing is more than merely seeing, it requires preparation: decide what type of observational data are needed, observe the activities, record the observation, and analyze the data.

Observational techniques are broadly categorized into two types: structured observation which focuses on specific behaviors or objects and tries to tally their characteristics, e.g. type, number, currency, or frequency; and impressionistic observation which is more open ended and less focused on individual behaviors or specific objects or conditions. What matters most is to develop a systematic means of recording all observations.

Less structure is appropriate where the observational objectives are less precise. For example, a visit an on-farm testing site calls for some impressionistic observation.

Because observation notes are often sketchy, team members should take time immediately after the observation period to completely fill in the observation record. Abbreviations and codes are written out. Notes are made regarding items that need to be checked.

TABLE 9: DATA GATHERING METHODS

Strategies	Strengths	Weaknesses	Types of Information	Audiences/Parties
Document Analyses: reviews of official papers that constitute written record of country (econ., ed., demo., ag.)	1. Is economical 2. Provides good source of general background 3. Serves as best source of project objectives 4. Is unobstrusive 5. Requires no new data collection 6. May surface issues not noted by other means	1. Is potential inapplicability of information 2. Could be bias due to "selective survival" of documents, quality of record-keeping 3. May require research effort 4. May not be current 5. May be disorganized or unavailable 6. May contain unreliable statistics	1. Program materials 2. Research papers 3. Planning documents 4. Research proposals 5. Reports of other agencies working in country, previous reviews 6. Internat. data, gov't publications & host country data bank	1. Host country government and institutions 2. Donor agencies working in country 3. International organizations
Presentations: oral talks to group	1. Allows for discussion 2. Allows presentor to focus talk 3. Provides communication vehiclefor gov. officials with limited time 4. Efficient communication of core material to team	1. May be inefficient use of team time 2. Could be bias presentation of information 3. May not focus on key issues 4. Leaves limited time for indepth questioning 5. Series of presenters are often highly redundant	1. Policy, organizational strategies & major issues 2. Organization structure & objectives 3. General or specific operations 4. Program & project status	1. High government officials 2. Administrators 3. Research staff
Observations: organized reviews and analyses of behaviors	1. Provides eyewitness accounts 2. Allows comparison of words and deeds 3. Is a sensitive technique with trained observer 4. Occurs in natural setting 5. May surface issues not noted by other means	1. May cause artificial conditions 2. Is inadequate sampling of observed events, especially when access is limited due to language and protocol 3. Results in frequent observer unreliability 4. Requires a lot of time and personnel costs	1. Type, condition, & currency of labs, facilities, equip, & supplies 2. Staff performance & relations 3. Extension & field activities	1. Visual inspection 2. Staff activities & interaction 3. Extension staff and farmers
Interviews: series of oral questions	1. Allows indepth & free responses 2. Is flexible & responsive to individual situations. 3. Permits follow-up questions 4. Provides setting to explore sensitive issues 5. Insures high response rate 6. Provides information on review to respondent	1. Is costly in time & personnel 2. Requires skilled interviewers 3. Is difficult to summarize 4. Is open to situational or interviewer bias 5. Is difficult to compare results of interviews by different interviewers	1. Policy, organizational strategies, & key issues. 2. Research unit approaches 3. Staff knowledge, attitudes 4. Possible recommendations	1. Policymakers 2. Administrators 3. Program staff 4. Farmers

The actual observation phase is always heavily context dependent. What is observed will be affected by a variety of factors, including whether the team member is observing alone or as part of a group; whether the visit was announced or unexpected; and whether the people are accustomed to being visited. Each of these cases places impediments on the observer which must be overcome (e.g., bias due to changed behavior, time constraints, physical limitations). Bias, the effects of behavior change when the people know they are being observed, may be the most serious threat to observational validity.

Like interviewers, observers can influence what people say and do. Some people become inhibited while others put on a performance. As a consequence, the atypical rather than the usual may be observed. Fortunately, the "observer effect" usually does not last long and experienced observers can reduce its severity and duration. Among the methods to overcome this include: unobstrusive observation, participant observation (joining the group one is observing), and being such a constant observer that the effect disappears.

2. Interviews

Planned interviews are better interviews. They are more comprehensive, require less time, are more enjoyable for interviewee and interviewer alike, and present fewer risks. After identifying what information is needed and who is likely to have that information and selecting an interview as an appropriate data collection method, there is still a lot to be done in preparing for interviews. There remains: design of interview guides, briefing of interviewers, selection of interview approaches and formats, as well as general preparation for ad hoc interviews and informal discussions. Each is described and briefly discussed in the following subsections.

But first a word about field tests. Interview field tests by tradition are used to work out the rough spots in new interview guides. A limited number of interviews are conducted during the field test in a real setting with one interviewee and two interviewers, one serving as an actual interviewer and the other as an observer of the interview process.

As a potentially excellent information source, interviews are used extensively in evaluations. Interviews provide an opportunity to both gather facts and identify issues, but they usually involve a large amount of time and they are one of the most costly data collection methods.

Interview guides represent the team's best thinking of what should be asked of the different categories of interviewees. Each interview guide is designed with a particular set of objectives in mind. Therefore, the list of questions or guide should be comprehensive but not redundant. Interview guides are developed following several general principles. Although contents vary, every guide has a number of features in common.

They:
* Are appropriate for the particular interviewee.
* Have easy and uncontroversial first questions.
* Have a logical order to questions.
* Have clear concise questions.
* Include questions that elicit measurable information.
* Have no leading questions.
* Use a limited number of open-ended questions (limited to instances where new ideas are sought).

* Include only questions that the respondent likely has knowledge about (Interviewees will often guess if they don't know the information).
* Require only one decision per question.
* Have questions that can be interpreted in only one way.
* Are as brief and simple as possible.
* Have a format that is an easy to read.

In short, the interview guide is a framework for eliciting responses. The interviewer sets the framework but allows the interviewee ample freedom to respond within the framework.

* Since you are interviewing to elicit information, most experts agree that you should only be doing 15-20% of the talking.

* Unless it is on the guide, don't ask questions that can be answered simply with a yes or no. Try to draw out more informative responses.

* Unless you are conducting a group interview, try to interview in private. This arrangement puts the interviewee at ease, portrays a feeling that you think he's important, and doesn't break trains of thought.
  Indicate how long the interview will last (approximately). Ask if it would be possible to hold telephone calls or other potential interruptions.

* Do not be afraid to rely on your own body language--facial expressions, voice modulations, and gestures. These nonverbal mannerisms can put the interviewee at ease and keep him/her talking. Periods of silence can be helpful, allowing the respondent a chance to relax and gather his thoughts to complete an answer. Silences can also push an individual to complete an answer that he was trying to avoid.

* Be patient, if an interviewee wanders or can't decide, summarize where you are and repeat the question.

* Begin the interview by establishing your framework: who you are, what you're doing, why this interviewee, and how information will be used. Assure interviewee of confidentiality.

* Close the interview by thanking the interviewee and taking time to answer any questions or handle concerns.

All interviews have a general flow to them. The team members start by identifying who they are, why the evaluation is being conducted, and why the interviewee has been asked to participate. It is best to begin with a few simple questions on the position and background of the person interviewed. Then, the interviewer moves into some topic that the interviewee can respond to easily, moving progressively into more difficult topics. Questions are grouped around common themes instead of jumping from topic to topic. When shifting topics, transitions are used so it's clear to the interviewee that another (new) topic is under discussion. The team member tries to use different types of questions such as close-ended, open-ended, and probes - moving generally from close-ended to open-ended to probes. Questions which indictes which answer is "best" should be avoided. The interviewee should always be given ample time to ask questions and supply his own remarks.

Three general types of approaches can be used in conducting interviews. Each has its own advantages.

A non-directive interview approach is sometimes employed. This format provides the respondent with the most latitude and is used primarily when the interviewee has the needed information but might have difficulty communicating it. The interviewer provides a warm supportive environment to help the interviewee feel comfortable and keep talking. General questions are raised, probes are used very infrequently, whereas paraphrasing of the responses is utilized whenever appropriate. Silences in conversations are handled in a related manner with reassuring facial and body gestures. This is close to an informal exchange of information, but often ends up providing insufficient information.

A structured interview approach is very different from the nondirective approach. A specific set of interview questions are carefully worded and sequences are planned to insure that each interviewee hears the same wording of each question in the same order. This approach does minimize variation across interviewers, facilitate data collection and data analysis. However, it inhibits the interviewer from exploring new issues, as well as reducing flexibility needed to address individual differences and circumstances.

The most useful interview approach for evaluations is the focused interview. This approach pulls from the best of the two previously mentioned - a basic checklist is used to focus the topics for discussion and to standardize coverage, but it allows freedom and in-depth probing to meet the needs of the particular situation.

Interview formats -- their setting and structure -- have an impact on the overall interview quality. There are several types of interviewer-interviewee arrangements: one interviewer to one interviewee, one interviewer to a group of interviewees, or using two interviewers rather than one. Each of these has its own advantages and disadvantages. A frequently used interview format is the one-to-one meeting. This approach is used whenever an issue is either particularly important or sensitive, the interviewee requests privacy, or the team member believes the most productive exchange would take place in such a setting. In this situation, the interviewer has more control in directing the interview, can probe into issues, is able to guarantee confidentiality, and is probably more likely to get complete and honest responses. This type of interview could take place in the individual's work place or in a restaurant - both provide different advantages. However, this arrangement is not always feasible due to time constraints, and physical surroundings.

Many evaluation teams use a modified format, with two interviewers to one interviewee. This format provides a built-in check on information received from the interviewee when written notes are not possible. Experience has also shown that this arrangement allows sensitive issues to be raised knowing that if negative reactions appear, a good solution is readily available. Prior to the actual interview, the two interviewers meet to plan which one will raise the sensitive questions. The other team member is instructed to watch for negative reactions and introduce a new and neutral topic if appropriate. The rest of the interview would then largely be conducted by this person and steered away from any potentially sensitive areas. Unfortunately, this format is more costly of team members' time for planning, conducting, and recording.

The group interview where a host country person is surrounded by friends and/or colleagues usually creates a relaxed environment for discussion. The group interview could take place spontaneously or be arranged by an team member. It provides a good starting point for other interviews and is a rich source of data. In such instances, the interviewer has less direct control

over the flow of communication but can gain a wealth of information. A second member of the team helps record the key points in the discussion. The group interview situation provides a chance for the interviewer to see issues explored among individuals, hear different perspectives aired, and listen to a multitude of topics brought up. Often the team member conducting the interview will have to follow up some points raised in the group discussion with selected individuals later. Good procedure also calls for some private interviews with one or two members of the group who were not active participants in the discussion. The group interview format has the added advantage of showing the team's concern for gaining a complete understanding of events and may be appropriate to get inputs from a group of junior researchers.

Even though the team plans its itinerary, discusses interview formats, and arranges for data collection activities throughout each day, conversations and meetings also take place informally. Due to the frequency of the informal conversational interviews, team members become adept at conducting impromptu discussions and realizing the importance of keeping a record of these exchanges.

The exchanges usually take place in relaxed unstructured surroundings when the other person's main thoughts are not focused on being interviewed per se but rather on simply discussing current activities or immediate surroundings, for example, while travelling. Responses in these settings are often more spontaneous and candid than in the office or laboratory setting. These exchanges present opportunities to explore new issues, raise questions of clarification, talk with staff not on the interview schedule, and generally allow people in the agricultural research community of the host country to voice new concerns and offer new insights.

This permits the interviewer to be responsive to individual needs and unique situations. Questionning is usually centered around the immediate situation and elicits clear concrete examples from respondents. Unfortunately, these exchanges are seldom complete. Therefore, the benefit of these informal conversations is enhanced when follow-up conversations building on the earlier talks and materials gathered elsewhere can be arranged.

Due to the large number of unplanned interviews as well as to update planned interviews, team members try to keep running lists of questions and key issues for further exploration. In this manner, they are always current about important issues to the mission that need answering.

3. Keeping Records

The goal in recording information is to have a summary record of the presentation, observation, or interview that is accurate, complete and useful. An experienced interviewer and observer knows the limitations of memory. Impressions that seem so clear during and immediately after an interview or observation are substantially weaker at the end of the day, and fainter yet at the end of the week. Without adequate records, the remarks from one interview can "drift" into another; the sites at one station blur into other observations. Confusion or uncertainty of sources is the result. For these and other reasons, good interviewers take thorough notes during the interview whenever possible and review the notes for completeness immediately after the interview. A second review of the notes at the end of the day is standard procedure. For those times when taking notes during the interview is not possible or acceptable, writing up the interview as soon afterwards as possible is crucial.

3.1. Guides and Checklists

The team uses several different means of recording, observational and interview data. The most specific procedure is the use of a guide. The guide includes a set of questions or items with a specified recording method for each of the questions - e.g., presence/absence, frequency, rating scales, or key words. The guide simplifies later analysis and enhances data consistency because of the standardization of questions and categories for responses. In addition to recording the observed activities, the team member notes the time of year, geographical region and other relevant facts or conditions. This additional information is important because the conditions under which the observations were made may be directly related to what is observed (e.g., activity in the fields is affected by the fact that it is the rainy season).

Checklists are similar to guides but less standardized in that the questions are not completely specified. The response categories usually are not provided on the checklist. The team members have a list of questions on the checklist but are given complete discretion in choosing which to ask. With a checklist, questions are on the instrument primarily to describe areas of inquiry. The checklist is used in reviews where there is less interest in compiling or combining the responses from different interviews. As with interview guides, however, the team member is expected to describe, on the interview checklist, the conditions under which the interview was conducted or the observations were made.

The team's observation activities cover a broad range, anywhere from the observation of laboratories where a guide or checklist might be ideal to an individual farmer's plot of land where narrative field notes might be more suitable. Guides can also be used for recording relationships between individuals. For example, how a station director relates to researchers or how an extension agent communicates with farmers.

Both guides and checklists are regarded as standardized forms. As such, they have several important attributes. They:

* Help the interviewer remember what to cover and state the questions in a manner that has been proven to be successful in previous missions.
* Help the observer remember key things to observe.
* May help the interviewer and observer with preprinted categories of responses that can be simply checked or circled.
* Help the team members identify topics or areas where information is needed.
* Make it much easier to combine information from different interviews and observations, even those taken by different members of the team.
* Make it much easier for other members of the team to keep informed of everyone's findings.

3.2. Field Notes

Field notes are not standardized. The team member merely takes notes, often just key phrases, sketches, or a personal shorthand perhaps in a small inconspicuous notebook. Although the field notes provide limited opportunity to combine results of different interviews and observations, they may be the most convenient and practical in some situations. This includes the situation where the people being interviewed or observed would be distracted or inhibited by the use of a guide or checklist. It also includes a variety of other situations where the team member is either engaged in a discussion of topics not adequately covered by the guide or checklist or engaged in an activity that makes the use of a guide or checklist impractical (e.g., in a car).

The field notes need to be written up in a form that other members of the team can understand as soon after the interview or observation as reasonably possible. Moreover, even prior to the full write-up of the field notes, it is very important for the team member to review the notes for completeness of at least the major points that need to be recorded. Hopefully, the initial review will occur immediately after the discussion and observation.

4. Data Analysis

Data analysis is a natural extension of planning and data collection. During the analysis phase of the evaluation, the information collected is further shaped into a form that is usable for discussing findings, developing recommendations and writing the final report.

Quite clearly the type of analysis chosen is dependent upon the particular issues being examined and upon the nature of the data itself. The data are organized around the most important topics of the mission as described in the terms of reference and review objectives. The analysis is designed to bring out the key issues identified during the mission.

The most frequently used method of analysis for information obtained through observations and interviews is a technique called content analysis. It is a hybrid of quantitative and qualitative methods. Content analysis involves developing suitable categories for classifying responses, recording the frequency of responses in each category, selecting verbatim comments which seem typical or highlight special concerns. Content analysis is important for two main reasons: first, so that one or two particularly well-written comments do not overwhelm the actual trend: and second, so that as much useful information as possible is obtained from the collected data.

Special attention is given to observational data which are generally analyzed in light of the other data. The analyst asks a number of questions. Do the observations confirm other data from presentations, interviews, and documents? If not, were the observations representative? Do the observations indicate a need for more investigation into any particular area? In cases where structured observations were used, the observations can be tallied and analyzed like questionnaire responses.

Much of the analysis of presentation, interview, and observational information consists of simple tabulations with sufficient indexing information to retrace the source of comments or observations. Common sense is the best rule:

* Do focus on the most important findings.
* Do not draw strong inferences from one or two comments.
* Do give appropriate weight to frequently or infrequently made comments.
* Do indicate where there is diversity in the data - eg. diversity in farm practices, diversity in the quality of agricultural research.
* Do not make statistical assertions (i.e., "17 out of 20 farmers...") if the cases were not selected in such a way that they can be used as a representative sample.
* Do describe areas where follow-up activities might be warranted.

Subject matter expertise also plays a strong but subtle role in the analysis of data. As noted previously, expertise has its influence on both planning and conducting of data-gathering activities. During the analysis phase, too, subject-matter knowledge influences fundamental decisions about what information is most important and what, possibly disparate information, needs to be considered together in addressing key issues.

VI USING EVALUATIONS TO REASSESS RESEARCH AND DEVELOPMENT POLICIES: A Retrospective Look at Rice Programs in the Ivory Coast

Chapter III demonstrated the advantages of integrating evaluation plans into program design since the very objectives of the program are likely to be better thought out and more realistic. This chapter illustrates the way in which evaluation findings can lead to revisions of research and development objectives and even to changes in policies in support of those new objectives. The "logframe" (logical framework) for the design and evaluation of development projects is a potentially powerful tool for planners especially when applied flexibly to serve the needs of the planner and not simply applied mechanistically to satisfy some bureaucratic need for standardization of project design or evaluation format. In this chapter we apply the logframe to look at the evolution of Ivoirian policy towards the rice sector. The fact that this is a retrospective look at Ivoirian rice policies gives us the benefit of hindsight and allows us to identify the particular points in the strategy where execution of the program was efficient but the basic hypotheses proved to be incorrect and vice-versa. The important point is that at several points in the evolution of the development program, planners were able to identify the problems and reorient the policies because they had implicit goals (appropriate evaluation targets) against which they could meaure their progress. Research priorities were then modified accordingly.

a. The Problem

The Ivory Coast came to Independence in 1960 in the middle of an export boom led by strong coffee and cocoa exports. Following Independence it experienced a two-decade period of rapid economic growth which averaged 8% per annum over the period 1960-1980. The Ivory Coast's rapid growth was aided by its close ties to France, its special role within the franc zone, a liberal investment policy and its conscious efforts to maintain the agricultural sector as an engine of growth. The rapid economic growth was accompanied by rising per capita incomes and an increasing degree of urbanization, factors which contributed to a rise in the per capita consumption of rice. Local production did not keep pace with the rise in consumption leading to large-scale imports which were seen as an unnecessary charge on the balance of payments. Consequently, the government declared a "rice crusade" beginning in the mid-1960's to increase domestic production and to bring down imports of rice. The crusade went through various phases as planners confronted unanticipated problems in realizing their objectives. The logframe becomes a useful tool in separating problems caused by the choice of research and development strategy from problems experienced in the execution of the chosen strategy because of other factors.

b. The Logframe

The logical framework provides a tool for project design and evaluation which:

1. defines the project inputs, outputs, purpose and higher goals in objectively verifiable terms;

2. hypothesizes the causal relationship between the inputs, output, purpose, and goal;

3. articulates the assumptions (external influences and factors) which will affect the causal lnkages; and

4. defines the indicators which will permit subsequent measurement or verification of achievement of the defined outputs, purposes, and goals.

The logframe implies that there is a causal linkage which runs from inputs to outputs to purposes and to higher goals. The management of a project is principally responsible for transforming inputs into outputs; whether or not the achievement of those outputs leads to the achievement of the projects purposes depends on the validity of the strategy of the project. The logframe implies that each lower level in the hierarchy of causation is necessary for the attainment of the higher level; if the assumptions about the way in which constraining factors are overcome prove to be correct, each lower level in the hierarchy should prove to be sufficient for the attainment of the higher level.

The logframe matrix, Table 11 (on page 50) is a four-by-four matrix with "Inputs", "Outputs", "Purpose" and "Goal" as rows going from bottom to top in the hierarchy of causation. The columns from left to right are the "Narrative Summary" (of the given input, output, purpose, or goal), "Objectively Quantifiable Indicators" (of achievement for those elements), "Means of Verification", and the "Assumptions" which must be made about other factors, often outside the control of the project, to ensure achievement.

c. The Rice Crusade 1967-70

The first phase of the rice crusade was a "crash program" designed to fill the gap between domestic consumption which had growth rapidly over the first seven years since Independence. Table 10 presents data indicating the nature of the problem preoccupying Ivoirian officials at the time:

Table 10: Production and Consumption of Paddy and Imports of Milled Rice

Item	1960	1965	1966	1967	1968	1969	1970
Domestic Production	160.0	250.0	275.6	340.0	365.0	303.0	316.0
Domestic Consumption	220.0	364.0	412.0	380.0	444.0	394.0	460.0
Imports (milled rice)	42.0	58.8	82.8	43.0	46.0	56.0	78.8

Note: One ton of milled rice = 1.6 tons of paddy equivalent.

Table 11: Logframe for Rice Program 1966-70

Narrative Summary	Verifiable Indicators	Means of Verification	Important Assumptions
Goal:			
- reduce rice imports to zero by 1975	- imports of rice	- balance of payments statistics - import licences	- consumption per capita estimates correct - urbanization and income assumptions correct
Purpose:			
- transfer market-surplus of 100.000T of milled rice to urban areas	- sales by mills mills of polished rice	- SATMACI milling statistics - chamber of commerce sales	- increase in paddy production - milling yield 0.60 - merchants willing to handle local rice
Outputs:			
- increase in paddy production of 160.000 T sold to mills	- yield and area planted	- yield surveys - extension reports	- extension efficient in transmitting packages and services - package works under small farm conditions - guaranteed producer price attractive - government buys all paddy offered
Inputs:			
- improved seed reaches farmers	- quantity of seed distributed	- SATMACI statistics	- Research and multiplication services coordinated
- supplies and services	- fertilizer/ credit	- SATMACI statistics	- SATMACI organizes distribution and supervises repayment from
- extension reaches x % of farmers	- Ratio of Moniteurs to area planted	- SATMACI personnel records	- SATMACI create trainning center, recruit job, specialize by region
- decision by government to set producer price+guarantee	- producer price	- interviews with farmers	- price is attractive and actually paid
- organization in parastatal efficient	- creation of SATMACI-Riz	- statistics re: training of moniteurs, seed, fertilizer sales	- parastatal form more efficient than Ministry

When rice imports reached 83,000 metric tons in 1966, double what they had been at Independence, the government declared a "crusade" to eliminate rice imports by 1975. The elimination of imports involved not only substitution of existing imports but also provision for foreseeable increases in demand. The rate at which the population would change its tastes from roots and tubers to cereals was not known at the time although it was well-known that consumption of cereals was much higher in urban areas than in rural areas.

Table 11 presents the logframe for the "crash program" 1967-70. The strategy for the short run concentrated on the 90% of rice producers (accounting for 67% of the total domestic production) who were in upland rice. There were several advantages in this strategy: it concentrated on the mass of producers already familiar with rice; it required little change in their techniques, IRAT (the French "Institute for Research in Tropical Agronomy") was relatively advanced in its research on rainfed rice and the country had a seed multiplication capability. The program called for provision of improved seeds, fertilizer, and a very close supervision by extension "monitors". Let us look at the elements that were brought to bear on the problem. A successful program required research results, organization, and supporting policies.

IRAT had already produced an improved variety of rice, Morobereken, which was ready for extension on a wide scale. It was a rustic variety that could withstand the vagaries of weather, it was resistant to seed shedding, and resistant to local strains of pyriculariosis. It had the disadvantage of a relatively long cycle (145-150 days) and presented problems when planting dates were delayed by late rains and was subject to lodging. However, it was a significant improvement over traditional varieties. Researchers had obtained an average of 2.1 Tons per hectare under rainfed conditions and planners believed that farmers would achieve 1.8 Tons with fertilizer and supervision. Through exchanges with Brazil, Zaire, Senegal and Upper Volta, researchers introduced other varieties with desirable characteristics suitable for particular regions of the Ivory Coast.

In terms of organization, the Ivory Coast had positive experiences with crop-specific programs entrusted to autonomous parastatal agencies. Consequently, the government gave the rice program to SATMACI, the Company for Technical Assistance for the Modernization of Agriculture in the Ivory Coast, the parastatal body which had successfully carried out the cocoa rehabilitation campaign. SATMACI adopted an extension strategy which implicitly technological change as external to both the farmer and the extension worker. Scientists developed the improved packages; low level moniteurs (extension workers) were taught how to use the improved packages and carry out demonstration plots (and little else), and the moniteurs then saturated the countryside in the planned ratio of 1 worker for every 100 hectares. It should be noted that at this stage the Ivory Coast was still building up its educational system from a relatively low base. Only a small percentage of students went on to secondary schools and few of the fortunate few would work in the countryside. Thus, the decision to carry out extension by large numbers of workers at low levels of training was imposed by circumstances. Given the crash nature of the program and the level of the extension workers themselves, it was felt that the role of the moniteur was to encourage farmers by demonstration, force the farmers to respect planting dates and certain improved techniques, and to make sure that credit and seed reached the farmers on

time. SATMACI did not have a supply of trained agriculturalists from which to recruit these moniteurs. It recruited from among primary school leavers with farming backgrounds and trained them in 6 month programs in special centers created for the rice program.

SATMACI provided through its supply centers the improved seeds, fertilizer, and equipment that farmers needed and the government built rice mills in the important rice-producing regions to mill and store the paddy that farmers were expected to produce. A guaranteed producer price of 18 CFA/kg of paddy was established throughout the country.

The success of SATMACI-RIZ in bringing together all these inputs necesary for the execution of the crash program is demonstrated in Table 3 below:

Table 12: Distribution of New Inputs by SATMACI-RIZ

Input	1965	1966	1967	1968	1969	1970
No. of Moniteurs	80	193	666	694	n.a.	n.a.
No. of Assistants	0	12	49	88	n.a.	93
Seed Distributed (metric tons)	99.5	259.2	595.2	1162.2	n.a.	n.a.
Fertilizer	182.5	544.6	662.8	1125.8	1463.0	1643.0

The crash program was carried out efficiently by SATMACI. It received the improved breeder seed from IRAT and multiplied it in sufficient quantity to meet the needs of the regional directorates who had their moniteurs with the techniques that research had provided for each region. There was an increase in production, however this increase was smaller than the increase in consumption, variable from year to year, and was not transferred to the commercial networks where it could substitute for the imports.

We can compare the goals of the SATMACI rice program with its expected achievements as follows:

Table 13 : SATMACI Rice Program: Expected versus Actual Achievements

Indicator	Expected Achievement	Actual Achievement
Irrigated Area in 1970	35.000	12.400
Domestic Paddy Production (T)	295.000	320.000
Imports of Milled Rice (T)	0	79.000
Tons of Paddy Milled	100.000	11.200
Yield at Milling	60%	53%
Cost of Rice (ex-factory)	35 CFA/kg	65 CFA/kg
Investment 1966-70(CFA)	3,670 billion	5.233 billion

We may say that the program was executed very efficiently but the strategy proved to be wrong, at least in terms of the ultimate goal of reducing imports, because there were factors which had not been taken into account adequately when the strategy was formulated. Several things occurred:

1. the farmers embraced the new seed, credit, fertilizer, and extension advice and did produce more paddy;

2. traditional merchants, the Dioula, offered prices for unthreshed paddy above the guaranteed producer price offered by the government-owned mills and thus succeeded in diverting paddy to their artisanal hullers for consumption in rural areas (where increased rice consumption substituted in part for traditional consumption of roots and tubers);

3. the government-owned mills failed to attract an adequate supply of paddy and consequently operated far below their breakeven point for efficient operation;

4. credit reimbursement barely reached 20% since farmers who sold their paddy to the Dioula did not have to go to the government mill where the reimbursement would be deducted from their outstanding debts; and

5. the rice coming out of the government mills was often of a lower quality than that which could be imported. As a result, local wholesalers were unwilling to market it. The government had to institute a system of tied sales (jumelage) whereby a wholesaler was required to purchase one ton of local rice for each ton of rice he was permitted to import.

It is clear from the figures in Table 13 that the rice program in the Ivory Coast achieved an increase in paddy production and invested heavily in creating the conditions for further increases. However, it did not capture a sufficient share of the increase to have an impact on imports or to operate its mills profitably. Moreover, SATMACI recognized that it would have to improve the quality of its milled rice to compete on the urban market. For a while the government attempted to fight the Dioula by licensing the purchase of paddy and banning the importation of spare parts for their motor-driven hullers, however, it eventually decided that the Dioula provided a service to a segment of the market not targetted by SATMACI and were not the root of the problem of lack of supply of paddy. They decided, therefore, to reorient the program in a way which made the target explicitly a reduction in imports and not simply an increase in production that was presumed to lead to a reduction of imports.

d. Reorientation under SODERIZ 1970-77

The reorientation of the rice program may have taken on new urgency because of the devaluation of the French Franc in late 1969, a devaluation which raised the CFA cost of imported rice at a time when the government was not in a position to subsidize the consumer. Under the reorientation of the rice program for the decade 1970-80, the government took actions which affected the organization of the development effort, the nature of supporting policies, and even the character of the producers targetted by these efforts. This called also for a reorientation of the research effort that supported the development strategy.

The shift in strategy called for:

1. concentration of extension efforts on farmers producing paddy for sale to the government mills;

2. particular attention to irrigated rice in the North and flood rice in the Northwest;

3. an increase in the guaranteed producer price;

4. investment by the government in the creation of irrigated perimeters and their colonization by rice farmers under contract to the mills;

5. the creation of a parastatal body with responsibility exclusively for the development of rice cultivation.

In 1970, the government created SODERIZ, the Company for the Development of Rice Cultivation in Ivory Coast, which then took over all buildings, factories, equipment and operating funds previously owned by SATMACI-Riz.

A conditioning input to the revised strategy was the fixing of a retail price for rice sufficiently high to permit SODERIZ to cover its milling costs after paying a farm gate price which was attractive to the farmers. The former guaranteed producer price had little effect on the supply of paddy reaching the "modern" commercial networks because it was too low. (It took 1.6 kg of paddy to make one kilogram of "riz cargo" selling for 45 CFA. If the farmer had sold his 1.6 kg of paddy to the mill he would have received only 32 CFA. Thus his wife, with her mortar, or the Dioula, with his huller, were both quite willing to capture the 13 CFA in value added per kilo which, over a day's labor, earned them the equivalent of the wage of an industrial worker.) During 1971 and 1972 the government benefitted from low import prices of rice and from subsidies from an "equalization fund" (Caisse de Péréquation) which collected taxes on imported rice to enable it to maintain the retail price at 55 CFA/kg. However, as soon as world prices began moving upwards it was forced to raise the retail price. Table 5 demonstrates how little room existed for encouragement of farmers when the retail price was maintained at 55 CFA/kg:

Table 14: Intermediate Costs from Paddy to Rice at Retail Level

Item	CFA/kg in 1969
RETAIL PRICE OF RICE	55.00
-retail margin/kg	2.70
-wholesale warehousing	0.20
-milling, grading, packing	6.20
-treatment, storage at factory	1.82
-transport to warehouse	2.15
-transport to collection point	3.00
-purchasing, handling	1.50
-impurities (stones, humidity)	1.00
-sacks for collection	0.83
TOTAL INTERMEDIATE COSTS	22.00
MAXIMUM FARM GATE PARITY PRICE FOR MILLED RICE	33.00
MAXIMUM FARM GATE PADDY PRICE (at 0.6 conversion)	20.00

In 1973, citing rising world prices, the government raised the producer price for first quality paddy to 28 CFA/kg and for second quality paddy from 20 to 23 CFA/kg. The retail price rose to 80 CFA/kg for milled rice.

The shift to irrigated production was aided by funding from the European Development Fund, the Coffee Diversification Fund and by technical assistance from Taiwan. The presumed advantages of full irrigation were the high yield potential and the protection that full irrigation offered from the vagaries of climate. (Putting ourselves in the planners position, we note that this was a period of drought in the Sahel which necessarily had an influence on the Ivory Coast). There were, however, certain disadvantages with the strategy. The Chinese semi-dwarf varieties (Taichung Native I and I Kong Pao) were found to perform well under rigidly controlled conditions but did not fare well under other conditions and were susceptible to disease (pyricularia oryzae) where water stress occurred. IR 5 and IR 8 became the basis for wet rice in the Ivory Coast. While these IRRI varieties are high yielding, and resist lodging they are sensitive to disease under stress conditions.

The problem of irrigation in the Ivory Coast, however, lies more with respect to the nature of the population into which it is being introduced than to the failure of researchers to produce the appropriate varieties. The problems were both social and technical. In the first place, the Ivoirians had little experience with irrigation and did not adopt it readily. The government, faced with its problem of imports, began to grant land to foreign Africans (mostly Malians and Upper Voltans) who

agreed to the contractual conditions of selling their paddy to the SODERIZ mills. This cession of land and concentration of investments for the benefit of foreigners created strong opposition among the local population. The problem was also technical: SODERIZ itself had to upgrade its extension staff before it was prepared to undertake the task of preparing the local population for full irrigation.

In addition to the very high investment cost of irrigation with full control of water, there is a management requirement which went beyond the stage of development of the farmers. Moreover, research still had not fully solved the problems of weeds, disease, and seed storage for irrigated varieties.

While the policy stressed the shift to irrigation, it is a fact that the increase in production that took place over the period 1970-77 was largely due to increases in output from the rainfed areas, which continued to receive distributions of seeds and fertilizer, and which responded to a large increase in the producer price. Table 6 presents the principal rice statistics for the period 1970-77.

Table 15 : Principal Rice Statistics 1970-77

Year	Area (1000ha)	Paddy Production (1000 tons)	Rice Imports (1000 tons)	Producer Price (CFA/kg)
1970	289	316	78.8	20
1971	282	385	97.3	23
1972	282	320	77.1	23
1973	290	335	147.9	23
1974	317	406	73.0	75
1975	390	496	1.6	75
1976	398	460	2.3	75
1977	409	477	147.5	75

It is clear from the above figures that the increase in paddy production which took place over the period was due to an increase in area planted to rice rather than to increases in yields that could only come from irrigation. The large increase in area came in 1974-77 when the increase in the guaranteed producer price to 75 CFA/kg encouraged farmers to increase their deliveries to the SODERIZ mills. In 1976, the government began to claim that its policies had eliminated rice imports except for "luxury" imports.

Unfortunately, the policy did not lead to a sustainable level of production for several reasons:

1. SODERIZ was both financially and physically unable to purchase, mill, and store all of the paddy that was delivered to it at the guaranteed producer price of 75 CFA/kg. In 1975, SODERIZ mills treated 100.000 tons of paddy which took them to their capacity. Once SODERIZ was no longer able to treat or store additional paddy it began to turn

away farmers who were then forced to sell their paddy on the traditional market at approximately 60% of the guaranteed producer price. The refusal formally took the form of refusing paddy with excessive humidity or too high a degree of impurities but the farmers understood the actual situation and have since proven to be skeptical of price promises.

2. The source of funding for the support price established for rice had traditionally been the Caisse de Péréquation which taxed imported rice and used the proceeds to support the price paid to domestic producers. When world prices were low and imports were significant this provided adequate resources. However, with rising world prices and declining imports both the tax wedge and the base to which it applied were becoming smaller. Since the Government was unwilling to see a permanent drain on is other resources by continuing to subsidize rice consumption to this extent, it raised prices to consumers.

3. Imports began rising again in 1977 as consumption continued to grow and farmers began to return to their traditional crops which they had abandoned for rice when its price was abnormally high. Farmers demonstrated once again that they have a high degree of responsiveness to prices incentives: they will produce more rice (and less maize, cotton, or yams) when the price of paddy rises relative to the price of those alternative crops. However, they also demonstrated that there is no theoretical or practical reason to believe that they will produce more of all crops if all prices rise at the same time and the rise in the price of paddy was shortly followed by pressure to raise cotton prices as well.

4. There were serious questions raised about the management of SODERIZ and of the whole group of parastatal organizations which had grown up since Independence. Abuses of the autonomy, mismanagement of funds and a general lack of coherence among activities led the Government to undertake a general investigation of the role of parastatal bodies.

d. The Search for a Strategy: 1978-84

By the mid 1970's both research and development were studying the need to change from a sectorial strategy to a regional strategy. It was recognized that successful sectorial programs soon ran up against constraints imposed by other crops in the farmer's system and thus a more integrated approach was recommended. On the research side, the Government began to talk about Ivoirianizing and regionalizing the French institutes working in the country: instead of maintaining separate crop-specific institutes the Ivory Coast wanted to integrate them into regional institutes. IRAT in Bouake would be joined with the livestock institute and the cotton institute to form an Institute of the Savannahs. This reorganization of research is still not complete since it involves several bilateral issues having to do with Franco-Ivoirian cooperation. At the same time, development efforts were regionalized. SODERIZ was abolished and its functions were divided among a series of parastatals with regional responsibilities. The cotton agency took over "development research and extension" in the North, the oilpalm agency took over these functions in areas where it was represented, SATMACI regained its role in the coffee and cocoa areas, while the Office for the Commercialization of Agricultural Products took over the operation of the rice mills.

This reorganization of research and development takes place against a background of a generally healthy agricultural sector with some worrisome tendencies in paddy production. The trend in the production of basic foodstuffs indicates a continued high responsiveness of farmers to market conditions and underlines the importance of stable marketing conditions if one is to carry out and monitor a cereals policy. Table 7 presents the gross production of principal food crops in thousands of tons:

Table 16: Gross Production of Principal Food Cropps (1000 T)

Crop	1978	1979	1980	1981	1982
Paddy	504	534	511	450	400
Maize	264	275	280	287	415
Millet	45	46	47	44	48
Sorghum	32	33	34	33	51
Yam	1984	2068	2068	2230	3050
Cassava	1056	1112	1112	1218	921
Plantain Banana	1123	1178	1178	1278	1356
Taro	274	287	287	304	340
Groundnut	---	---	---	56	67

The Development Plan for 1976-80 had called for the investment of 21 Billion CFA to develop 35.000 hectares with full or partial water control and to add 30.000 hectares of improved upland cultivation. On the basis of this investment, the Ivory Coast was forecasting the production of an additional 350.000 tons of paddy by 1980 which would leave only 65.000 tons of rice to be imported. Even the World Bank, citing the impressive responsiveness of farmers to improved prices in 1974 suggested a scaling down of the investment targets in irrigation and for more attention to improved cultivation of rice in stable rotations in the savannah areas.

From 1980 onwards, there was some uncertainty about the strategy to be adopted. The Minister of Agriculture attributed the faltering of the rice program to the continuing drought and the "poor management" of SODERIZ. The turning over of the program to parastatals without any previous experience in rice led to a period of marking time while they became familiar with the host of problems facing the sector. (While they might have been organized to provide extension to farmers they were certainly unfamiliar with the supply and marketing problems). It was estimated that in 1979 there were 24.000 hectare/cycles (offering a production potential of 84.000 tons of paddy) potentially available in the Ivory Coast but that only 40% of the capacity was utilized.

The Minister defined the new policy as follows:

> Our production rests practically 90% on rainfed rice and the dependence on rain poses very serious problems for us. One of the strategies envisaged since 1970 is therefore to produce under water control: irrigated rice. But there is a very elaborate technique to which the Ivoirians were not accustomed and which demands a certain discipline. The several technical difficulties which we knew (drought) added to those experienced by the parastatal extension organizations have meant that this production has stagnated and even fallen. The new policy of the Government is to continue with rainfed production but this time in rotation with, for example, cotton, soybean, or in association (with coffee) and above all to intensify production of irrigated rice in creating or recuperating the lowlands which were equipped five or ten years ago.

The relative stagnation of rainfed cultivation was attributed to difficulties experienced in reorganizing the extension programs and to the diasaffection on the part of the farmers created by uncertainties as to the market guaranteed by the government. In fact, in 1981 only 10% of the area under cultivation was supervised by extension workers. The target is to have supervision of 94.500 ha of rainfed cultivation and 50.500 ha of irrigated production (or 30% of the area cultivated) by 1985. Even if this goal is achieved, the country may still be obliged to import about 40% of its rice needs.

The Plan Paddy outlines the contribution of each type of rice cultivation to the reduction of imports over the period 1980-1990. Table 17 presents the targets and the requirements for achieving them for 1985 and 1990.

Table 17: The Plan Paddy (tons of paddy)

	1980	1985	1990
Global Production Objective	396.000	570.000	975.000
Coming from production trends	396.000	430.000	455.000
Operational Production Objective	--	140.000	520.000

The planners estimated the growth in consumption based on certain assumptions about growth of incomes, urbanization, and consumption per capita in rural and urban areas. Projecting the trend of output of paddy they derived an estimate of availability of paddy without major new efforts. The difference between demand and supply became their "Operational Production Objective". In terms of the logframe this would correspond to the row "outputs".

The planners then looked at the "strategic means" to achieve the additional output:

Table 17 (continued): The Plan Paddy

Strategic Means: Expected Production	1985	1990
-Irrigated Cultivation		
-recovery existing areas	39.000	39.000
-new developments	61.000	121.000
-Better extension and marketing in traditional rainfed areas	40.000	70.000
-Mechanized rainfed rice within modernization programs	(*)	290.000
of which: animal traction		(100.000)
tractorized cultivation		(190.000)

Physical Inputs Required:			
-New development of irrigated lowlands at 1.5 cycles/ha and 3.5 T/cycle	Hectares:	11.600	11.400
-Animal Traction (1.5 T/ha)	Hectares:		65.000
Tractorized Cultivation (3.0 T/ha)	Hectares:		57.000
Imports of Rice Required:	Tons:	250.000	200.000

(*) Any increases coming out of these programs will serve to reduce imports.

It is evident that the government is looking to a major effort in irrigated cultivation over the period 1980-85 and to new programs in mechanized cultivation over the period 1985-90 to reduce the rice deficit. Both of these orientations assume that the population is able to assimilate and participate in the new techniques proposed.

The Government proposes a number of supporting actions which are considered necessary to improve rice marketing:

1. the opening and maintenance of rural roads

2. increased emphasis on cooperatives to take advantage of the fixed marketing commissions for each function from collection at the farm gate to delivery to the mill;

3. increased capacity at the mills to buy and store paddy (which will raise the milling capacity from 100.000 T in 1980 to 280.000 T in 1985); and

4. improved diffusion of information about prices and markets to allow both Government and private firms to react to distortions in the market.

Since January 1984, the Government has set the price of paddy at 80 CFA/kg at the farm gate with additional commissions payable to farmers or cooperatives who collect and deliver their paddy to the mill (where the price is 95 CFA/kg). This may be considered an incentive price to encourage farmers back to rice and to help overcome an estimated 300.000 Tons of rice that are expected to be imported in 1984 as a result of the drought and continuing high consumption.

e. Implications for Agricultural Research

It is clear from the objectives outlined in Table 8 that the Government looks to lowland rainfed rice with some water retention facilities as the intermediate solution to the import problem. The technique is seen as an excellent introduction to intensive irrigated cultivation and involves smaller engineering investments and a lower level of management than full-scale irrigation. Researchers have argued that it is not necessary to look for a high yielding variety with a maximum potential but to find a variety with the necessary resistances (drought, flood, disease) which can be sown across a wide range of ecological conditions and still give stable yields. The current recommended varieties all have some properties which are undesirable. (Iguape Cateto originally from Brazil has a tendency to lodge and exhibits some seed shedding while IM 16 from Guinée has disease problems and is photoperiodic.)

It is quite clear that the research services have been able to respond to the different orientations in policy and bring usable varieties of upland rice, irrigated rice, and rainfed lowland rice into multiplication. However, given limited resources, the research services must be able to concentrate their effort on development of varieties which will be supported by realistic development policies. Rice is considered one of the important commodity programs in the Ivory Coast but it accounts for only 15% of the scientific manyears devoted to research on commodities. The ISNAR review of the Ivoirian system notes the following allocation of manyears

Table 18: Scientific Manyears Allocated to Principal Crop Programs 1983

Program	Total Number of Scientific Man-Years
Oilpalm	17.5
Coconut	8.5
Rubber	14.5
Coffee	10.0
Cocoa	9.5
Pineapple	9.0
Banana	8.0
Rice	14.0
Cotton	7.0
TOTAL LARGE PROGRAMS	98.0

The 14 scientists working on rice are, consequently, divided among programs aiming at improved upland rice for use in mechanized systems, flood rice, and irrigated rice. Undoubtedly each type of rice has its role to play in the short and long term strategy but 14 scientists are unlikely to produce the research results that a "crusade" would require. There appears to be a consensus on the need to develop a variety that offers stable yields in both pure rainfed conditions and in lowland cultivation with some retention of water.

Concluding Observations

In the preceding pages we have attempted to analyze the evolution of Ivoirian rice policies since the beginning of the rice "crusade" in 1965. The initial strategy based on extension and marketing efforts among traditional farmers in upland cultivation proved inadequate to the task. A reorientation of the program towards development of fully-irrigated cultivation ran up against a constraint imposed by finances, price policies, and the difficulties in spreading the method among Ivoirian farmers. Following a reorganization of the research and development structures, the Government is looking to the development of lowlands where some form of water retention provides an introduction to irrigation and lays the groundwork for future development of full-irrigation among the population. A parallel effort is called for in the development of upland rice in stable rotations with mechanized cultivation. Both of these strategies will call for investments and supporting research enabling the rice sector to advance its general level of technology beyond what has been known in Ivory Coast to date.

VII CENTRAL MONITORING AND EVALUATION

This reprint of a course note (CN-318, Nov. 1980) from the Economic Development Institute of IBRD describes how monitoring and evaluation procedures are organized in a few countries.

This paper is concerned with the role of central monitoring and evaluation[1] units in bringing about a more effective achievement of the development objectives of Third World Countries. About twenty such units have been established in the L.D.C.s, most with the objective of identifying and resolving difficulties experienced in planning and implementing development projects and programs. These units have been located in the Central Planning Agency, in the Ministry of Finance, or in other agencies of the central administration. Their procedures, their systems of gathering data, the analyses that they use, and the administrative impact that they have are different and varied. From this diversity, however, come a number of basic principles and practices which can be used to strengthen the overall development process, and to enhance the contribution that a central monitoring and evaluation unit can make.

The paper is based on an on-going study of the actual practices of monitoring and evaluation units in a number of Third World Countries. This study is not yet completed, but sufficient material has been gathered from which to develop some reasonably firm conclusions. The conclusions have been based on a review and analysis of the systems that are in operation, on an analysis of the administrative and managerial concepts underlying monitoring and evaluation, as well as on the many

1 For a definition and discussion of the terms "monitoring" and "evaluation", see Appendix A.

discussions with officials who operate these systems in the L.D.C.s and in the bilateral and multilateral agencies.

Central monitoring and evaluation units are not a new concept. They were first introduced in a few countries in Asia and in Latin America in 1961 or 1962, although India had started a central evaluation unit in the Planning Commission as early as 1952. Some of the early monitoring activities were conceived as a method of supervising the managerial performance of the implementing agencies, but the objectives of most units were largely as they are today i.e., to held identify and resolve problems that were occurring during the implementation of projects and programs. The original units were also conceived as systems to provide information for the visual display of progress achieved in national developmental efforts. This concept, which appeared to have been drawn from the "operations room" or "management information center" concepts then prevalent in private industry, has since been dropped. Only two countries of the eight surveyed to date, Malaysia and Sri Lanka, still have a display center.

Another change that has occurred in the last decade has been an increased emphasis on the role of evaluation in the development process. Although the early units were concerned with evaluating the methods and processes of implementation, more recent evaluation efforts have emphasized the assessment of the impact of the project components and the extent to which the project will achieve its objectives. This emphasis has been a significant one, but has raised problems of combining the monitoring and the evaluation functions of the units, and of determining how current evaluation techniques can be incorporated

into the activities of a central unit. The original terms used to describe the work of the units were "monitoring", "project control", or "project follow-up": activities that were concerned with checking progress towards project objectives, and identifying and resolving the problems that were hindering this achievement. Although such activities also encompassed the evaluation of whether the project would achieve its objectives or not, the increased emphasis over recent years on rural and poverty alleviation programs has introduced complexities into this evaluation process.

There has been, therefore, a widening in the operational objectives of some of the units. In addition, the variety of operating methods that are used to achieve these objectives raise questions as to which are the more effective approaches that could be adopted. This paper examines and attempts to answer such questions. It takes the results of the partially completed survey covering the eight countries*, and draws conclusions on the most effective approaches to deciding:

(1) The role of central monitoring units vis-a-vis ministry or project monitoring

2) The optimum objectives and functions of the central units. i.e. the extent to which the unit monitors progress and identifies problems that are occurring, or whether it also

---

Guatemala, Honduras, India, Indonesia, Jamaica, Malaysia, Peru and Sri Lanka. When completed, approximately 12 units will be covered by the study.

undertakes the task of evaluating the achievement of the intended social and economic impact of the project.

3) The types of projects and programs to be covered by the units. i.e. whether the unit covers all projects or only those key projects which have a significant impact on the overall rate of development.

4) The responsibilities of the unit in resolving implementation difficulties. i.e. whether the unit becomes involved in the problem solving process, or whether it simply reports difficulties to the responsible ministry or agency of government.

5) The optimum location of the unit in the administrative structure of the central government and its relationships with the operating ministries and the projects.

6) The methods of gathering, reporting and analysing data.

CENTRAL VS MINISTRY MONITORING

One factor that needs to be examined at an early stage is the relationship that the central monitoring and evaluation units have to any monitoring activities in the implementing ministry or in the project itself.

All project managers, of course, monitor their own projects and to the best of their abilities, evaluate the effectiveness of their activities. In the study, a few ministries also operated formal monitoring systems; very few had established formal evaluation systems. No duplication of evaluation activities, however, between the ministries and the central unit was observed on the study.

With the exception of two units that were monitoring overall plan implementation, however, all units in the study received their monitoring information directly from the project.* Their data gathering processes, therefore, generally bypassed the operating ministries. In the few cases where ministries operated monitoring systems, the same data were generally used by the ministry and the central unit.

Any action that was taken on monitoring information, however involved the responsible ministry, either through the unit reporting to the ministry on the progress and problems on its projects, and leaving any corrective action to the ministry, or the joint taking by the unit and the ministry of any corrective action that was necessary. In a few cases, major problems that were either common to many

---

* With some exceptions, this paper uses the term project to encompass both projects and programs. A project is defined as an activity with a finite completion date that forms part of a wider program. A program may have an end point (e.g. a program to develop self-sufficiency in food production) or it may be an on-going activity of government without any definable end point (e.g. a continuing rural subsidy program). A project, or a program, consists of a series of inputs and outputs in which the outputs from earlier series becomes the inputs to subsequent work effort, eventually culminating in, to use the terminology of USAID's Logical Framework, the achievement of project purpose and program goal.

ministries or which were a significant impediment to overall development were tackled by the monitoring unit on a unilateral basis.

Although it might seem that a three tier system of reporting, from the project to the ministry and thence to the central unit, would be an approach that would preserve the autonomy of the ministries, (and which also would provide a capability for an exception reporting component in the system) in no case was this three tier system developed. It would appear that there could be three reasons for the units having developed a direct project to central unit reporting framework:

(i) The relative speed and simplicity of a system that required one rather than two tiers of information flow.

(ii) The preservation of the ministries' autonomy through the way this information was handled after its receipt, rather than in through its reporting.

(iii) The advantages, as is discussed in a later section on the optimum location of a central unit, of having a unit that directly spans most of the major development activety in the country.

OBJECTIVES AND FUNCTIONS OF THE UNITS

A central monitoring and evaluation unit has as its overriding objective the improvement of the effectiveness and rate of development. The methods of operation that were used to achieve this objective, however, varied considerably. The various functions that the

units had adopted, none of which were mutually exclusive, and each of which varied in the methods that were used, were:

1 To monitor implementation, so as to improve the performance, cost and timeliness with which the objectives of the country's developmental projects are achieved. This function included a direct involvement in improving implementation in the case of some units, but not in the case of others.

2) To evaluate the social, economic or administrative effectiveness of on-going projects, in order to optimize the impact of current projects, or to improve the efficiency with which the current projects were being implemented (on-going evaluation).

3) To carry out evaluations on completed projects, in order to increase the effectiveness with which new projects are planned (ex-post evaluation).

4) To monitor performance, by sector, in the implementation of national development objectives (usually embodied in annual or five year plan targets). Such a monitoring process would be designed to strengthen the planning and plan implementation process, rather than the planning and implementation of individual projects.

5) To control capital budget expenditures by monitoring and compiling actual expenditures on a periodic basis. (This function is part of a broader monitoring function, but has been treated separately as it was a distinct and separate activity for one of the monitoring units).

Of the eight units surveyed to date:

- All eight had the objective of monitoring project implementation performance (but for two of the units, it was not a primary objective).

- Two countries, India and Malaysia, had separate evaluation units in the same central government agency as the monitoring unit, responsible for evaluating the social or economic impact of on-going and to a very small extent, completed projects. A further two units (Peru and Honduras) carried out limited on-going evaluation activities along with their monitoring activites.

- Two units, India and Sri Lanka, were working primarily towards the fourth objective of monitoring achievement against annual or five year targets. The two units, however, did not monitor broad indicators of development progress. In both these countries, much of the industrial and modern agricultural sectors were under public ownership, and the units monitored the performance of selected sectors of the economy by measuring the budget and production

achievements of government agencies and corporations. Both of the units also undertook some project monitoring.

- Only one unit in the survey (although a number of units not included in the survey are known to carry out this function) had the fifth objective of monitoring capital budget expenditures, In the survey unit, the function had been adopted as an attempt to strengthen the poor performance of the operational agencies in achieving capital budget targets.

The majority of the monitoring units in practice, therefore, had the objective of improving the implementation process. A small portion of them was concerned with evaluating administrative effectiveness and project impact. A more widespread adoption of the second, third and fourth objectives, however, can be achieved by adding data collection or survey research facilities into the system. They are not contradictory with the first objective. This possibility raises the question of whether units should include the additional functions of (i) a more widespread adoption of impact and process evluation practices, (ii) gathering and evaluating of information on completed projects that would help improve the planning projects and programs in the country, (iii) the monitoring of overall development progress through the adoption and measurement of various social and economic indicators and (iv) monitoring capital expenditures.

The answers to these questions will become clearer as we discuss the methods of operation, location and effectiveness of a central monitoring and evaluation unit. It may be mentioned now, however, that the additional functions belong, in practice, to a separate administrative activity in most Third World governments. They are essentially designed to strengthen the planning process. Implementation is a management function that uses different disciplines and skills. A unit that is monitoring implementation will therefore tend to use different skills then it would in evaluating project impact. These are possibly the reasons why the only two countries that had evaluation units kept them administratively separate from the units which monitored project implementation.

There are further factors at work which inhibit an integration of planning with implementation. One is the widespread practice under which planning units, either in the central government or in the operating ministries, have little responsibility for implementation. The possible reason why planning agencies do not concern themselves with implementation is that many planning agencies and units in Third World Countries, unfortunate as it may be, do not have the status, the direct influence, or the skills to examine problems of implementation, or to advocate successfully the policies and managerial decisions that would rectify implementation difficulties. Monitoring and evaluation activities tend to concentrate on improving implementation, therefore, and to exclude the evaluation activities which are more concerned with the planning process.

Nevertheless, it is possible, and highly desirable, to improve the integration of project and program planning with implementation.* Although the implications of this step would be to extend the practices of a number of the units, a more widespread inclusion of impact evaluation as one of the tasks of a monitoring and evaluation unit would certainly add to the overall effectiveness of such a unit.

In this paper we have adopted a pragmatic viewpoint. If the monitoring and evaluation units in practice have an objective of improving the implementation process, and if there are conceptual and administrative difficulties in extending this objective, then the primary focus of the paper must relate to the implementation process; i.e. to the determination of those practices which enable a central monitoring and evaluation unit to improve the implementation of projects and programs.

PROJECTS AND PROGRAMS COVERED BY THE UNITS

(a) Types of Program

In practice, most of the monitoring and evaluation units did not cover programs such as incentive or subsidy programs, etc., nor did

---

* In many of the newer projects, it is a vital step. The uncertain nature of the outcomes make a continuing assessment of results and a continued replanning of the project necessary for the successful management of the project implementation. See the author's paper on *Problems of Implementation*, Economic Development Institute, Washington D.C. 1980.

they often include social welfare programs such as maternal & child health, nutrition, family planning, etc. When the unit did include human resource development programs and related activities, it was often only the infrastructure aspects of the programs that were covered, and rarely was the economic or social effectiveness of these programs evaluated.

There are possibly three reasons for an incomplete coverage of this type of program:

1) Some monitoring units only covered the larger development programs, which tended to exclude social welfare projects;

2) The legacy of the monitoring developments of earlier years, in which the cost/time/performance aspects of managing infrastructure projects received the emphasis in the management literature, and which, in combination with the relative newness of programs aimed at improving the conditions of the poor in the Third World, has caused some uncertainity as to how the "new type" project should be monitored.

3) The relative ease with which the physical aspects of programs can be monitored. The outputs of social programs are difficult to monitor (although inputs are relatively simple), and usually require more elaborate social survey research methods.

There is, however, no reason why units should not monitor the more complex social development programs of the government. Some of these programs are likely, in fact, to have a more significant impact on

social well-being or on economic growth than many of the projects that are monitored. The problem, of course, lies in developing the systems, techniques and skills that will enable such programs to be effectively monitored and evaluated.

(b) Total coverage of projects versus coverage of key projects

An issue that has to be resolved by a government intending to establish a central monitoring unit is whether it wishes to cover all development projects in the country or whether it wants to monitor only those projects that require a large investment or which would appear to have significant developmental or political impact.

The advantages of limiting the coverage of the unit to key projects are:

(i) Usually a higher percentage of the development budget is spent on a few major projects than it is on the remaining multitude of smaller projects. Concentration on the few projects therefore is likely to provide control over a large portion of the current developmental effort.

(ii) The unit has no need to develop the sophisticated data collection and processing procedures, or to recruit the large number of highly skilled people that would be required if all projects were to be monitored. It is no coincidence that two out of the three systems in the survey that used electronic data processing were attempting to cover all projects. In these cases, the number of projects being monitored

ran into the thousands, and electronic data processing presented the only viable way of sorting the large amounts of data that were produced. Both these units, incidentally, acted primarily as data processors. Most of the problems that they identified were referred to the line ministries or to other agencies of government for resolution.

(iii) Monitoring of a smaller number of key projects and programs enables the unit to become more closely involved in each project, thus providing the opportunity for a more accurate analysis of the causes behind problems and, if desired, for contributing more effectively to the resolution of difficulties that may occur.

The advantages to be gained by limiting the coverage of the monitoring system to a few projects would certainly indicate that any country establishing a monitoring unit could develop a more effective system if it first covered only key projects. It could then use the experience gained in the process to determine the manner and the extent to which it wished to widen its monitoring coverage.

The advantage of a full coverage is that it enables the work of all government agencies to be reviewed. In addition, many small projects frequently have problems that are not experienced by larger projects. They are usually executed by the regional administrations of line ministries or by local authorities, they do not usually receive foreign assistance and, in their lack of visibility, are often starved of financial or skill resources. If they are not monitored centrally, then their common problems will take a longer time to emerge, and to be resolved.

Connected with decisions regarding the coverage of projects by the central monitoring and evaluation system, however, are questions relating to the types of problems which can be resolved, the responsibilities that a centralized agency can exercise in solving problems that arise in the executive departments of government, and the most effective location in the central government of a monitoring unit under the various operating conditions that could arise. These issues are discussed in the following sections.

THE RESPONSIBILITIES OF THE UNIT IN RESOLVING IMPLEMENTATION PROBLEMS

The objective of any investment made in a development project or program is to achieve a level of developmental progress greater than would be obtained by alternative uses of the invested resources.

The achievement of this objective can be negated in one, or both, of only two ways:

1) The inputs that went in to the project or program did not go in as planned. Either they were sufficiently late to delay the project, or they cost too much, or their performance as an input was below expected levels. The benefits that were derived, or the costs that were incurred were, therefore, not as planned.

2) The mix of inputs went in as planned, but they did not or could not, cause the expected output. This difficulty arises because the knowledge of expected input-output relationships is, during the

planning phase, incomplete or faulty. It is more likely to occur in programs that depend on the responses of individuals to developmental inputs. Infrastructure projects are the result of an engineering design and, although the end result does not always conform with the plan, there is less liklihood of poor or incomplete project design (or a design based on incomplete knowledge) being the reason for inadequate output performance.

An example of difficulties of both types (and of the entwined nature of monitoring and evaluation) that could be experienced in practice are the assumption that a new seeds distribution program would produce certain yields. The yields may not be achieved, however, because the planning assumptions on expected yields were wrong, or because the seeds were delivered too late, or cost too much to be acceptable to the farmer, etc.

Monitoring and evaluation should pick up any of these issues. It may require simple observation for the project manager to ascertain the cause of the problem, or it may require an elaborate survey research study.

The cause of a project's difficulties could also arise in one or more of several agencies of government. If we take as an example the seeds distribution program cited above, the poor yields could arise because:

- the site management responsible for implementation did not distribute the seeds to the cooperatives in time for the farmers to use them,

- the the purchasing officer at the central office didn't order the seeds in time,

- other agencies within the overall government administrative structure, (such as the seed supplying agency, or the financing agency, etc.) did not contribute as planned,

- planning bodies, either in the implementing ministry or in the central planning agency did not have sufficient information on which to plan accurately. For example, the assumption on expected yields may have been too optimistic given the type of seed, the soil conditions, the proposed time of planting, the acceptance of the farmers, etc. In this case, it was a planning issue that caused implementation problems.

Poor achievement can also of course, arise by pure chance (i.e. weather, or other unforseen impact on project outcome). Wherever the problem may have arisen, however, readjustment of the program may affect one or more of several agencies of government (i.e. the project may require different inputs, different schedules, etc, which may in turn have implications on budget and manpower allocations and the workload of any of the contributing agencies that are mentioned above).

Under these conditions, the question arises of the most effective location in the administrative structure for the monitoring or the evaluation of this project. A corollary is whether a central unit

should have a responsibility for correcting problems on the project when it has no direct authority. A central unit would certainly appear to intrude on the responsibility that ministers have for the efficient conduct of their ministries, or even on the managerial responsibility that the project manager has for the successful implementation of his project.

Centralization and Decentralization

Most writers on management and administration endorse a decentralized approach to decision-making. The reasons are that the need for the decision is usually identified more quickly, the decision itself is taken more readily and with a greater knowledge of its consequences, and it will be implemented with closer supervision and feedback.

These theories argue for monitoring and evaluation activities being carried out as close to the project or program as possible. However, these statements do not preclude a centralized monitoring function, provided that the results of its monitoring and evaluation are made available to the responsible agencies, and that any executive decisions are taken by these agencies.

There are, in fact, several reasons for centralizing the function to the maximum extent possible:

1) Managerial skills in monitoring and evaluation are in short supply in the administrations of most countries, but particularly in Third World Countries. There are advantages, therefore, in centralizing the necessary skills within one group. This statement applies particularly to the more complex evaluation activities which require a high degree of training and experience in research methodology.

2) Monitoring and evaluation units can provide a service to the implementing agencies, by obtaining monitoring information directly from the projects and programs, and providing this information to the operating ministries. The system can then be routine and standardised. When operated in this fashion (as was the case with some of the units in the survey), the monitoring and evaluation units would be taking no decisions which impinge on the operating ministries.

3) Many problems repeat themselves on project after project. A central unit is in a position to identify recurring problems more quickly, and to identify the steps needed to resolve them.

4) Many problems are not caused by the on-site management of the project or by their supervising ministries, but by the actions of the contributing agencies of government. In fact, much of the scheduling and supply of skilled manpower, funds, equipment and of supplies for a project lie outside the responsibility of the operating ministry. Although the implementing ministry should, and frequently does, appoint a project manager who has the responsibility for coordinating all contributors, such a project manager often has little influence outside

his own ministry. In any case, the best he can do is push for his own project. He frequently cannot even identify (let alone correct) any underlying problems or system deficiencies that may be causing problems in other projects as well as his own.

In summary, it would appear that a central monitoring and evaluation unit it could have a role in identifying and helping resolve problems that occur on projects that have major developmental impact, or on major problems that cut across many ministries. It would have no role on problems that are the responsibility of the project manager to resolve, and need not act in a supervisory or auditing capacity with respect to the operating ministries. It could provide various contributions to implementation problem solving, however, as follows:

(i) To supply the ministries with information on the progress and problems within each project under their direct responsibility. It would be the ministries' function to resolve the problems.

(ii) To identify project specific shortfalls in the supply of the inputs expected from the contributing agencies. The monitoring unit could confine its activities to reporting the shortfall, or to identifying further the reasons behind the shortfall. All of the units surveyed that did try to resolve shortfalls did so in cooperation with the implementing ministry.

(iii) To identify (and possibly help resolve) common problems that are occurring across several projects. These common problems may be simple shortages, or they may lie deep in the administrative systems of the central government. Again, most units in the survey that undertook to resolve major problems did so in conjunction with the concerned ministries. One exception, Peru, did undertake major administrative issues, but this unit was considered as a major management improvement unit of the government, and only undertook such tasks at the request of Cabinet.

(iv) To resolve, or help resolve difficulties that are occurring on the major developmental projects and programs in the country. Of the two units in the survey that undertook this work, both did so only on decisions made at Cabinet level.

In fact, most of the units that were monitoring project performance also reported problems to a higher political level (to Cabinet, to the Chief Minister, or to his deputy). The extent to which this information was used varied considerably. In addition, if the information was used as a check of managerial performance on individual projects, as opposed to a method of identifying and resolving common problems, the units tended to be regarded as auditers instead of being units designed to assist the staff of the implementing ministries.

The Resolution of Problems

As noted earlier, the functions of a unit may be confined to passing on the problems and progress information that is reported to it. In many cases, however, such reporting could be erroneous or even misleading. Poor contractor or contributor performance, for instance, may be reported by the site manager simply as a late delivery of supplies or equipment, but the underlying cause of these late deliveries (or excessive costs, or otherwise inadequate performance) may lie in the procedures used to select contractors, in the Ministry of Finance's cash management and payment systems, in the methods of contractor or contributor scheduling and control, etc. It is obvious that a monitoring unit would be more effective if it isolated those underlying factors in the administrative system that are causing project difficulties. To do so, however, it needs to have some investigative authority, and not to rely solely on reports from the managers of development projects. With such authority, and with the analytical capabilities that the unit must possess, a unit is placed in a strong position to identify the best possible solution to the problems that are occurring.

These statements are not to suggest that the unit in any way acts as an "auditor" or "policeman" with regard to the ministries or projects. Although such a role was a strong reason behind the formation of many of the early monitoring systems, (and which still lingers on in some current units), it has also been a reason for the rejection of outside monitoring and evaluation activities by many operating managers. Most of the units now go to great lengths to avoid

interference in the operating ministries' activities, handling the problem solving component of their activities in one of several ways.

(i) By reporting only progress and problem information to the responsible ministry. One unit, Indonesia, has a further capability in its system to report back to the project manager when the problem is resolved. This component was designed to demonstrate to the project manager that the system worked to his benefit. Guatemala had a similar capability, although both units also incorporated a follow-up or reminder component to the Ministry when the problem is not solved.

(ii) Establishing a joint ministry/monitoring unit review committee (Honduras). All corrective action identified by the monitoring system is decided by this committee.

(iii) Reporting to a committee of Cabinet of chief ministers, one of whom is the implementing minister (Malaysia directly, Peru via the units own minister). Any problem solving activity is assigned by that committee.

It is for the reasons that a simple reporting of performance does not always ensure an effective monitoring of projects, however, that a wide ranging responsibility for the unit, if sensibly used, will ensure a more effective monitoring system. The extension of a unit's responsibility to identifying those causes of implementation difficulties that range across the administrative structure, together with the involvement of the unit in seeking answers to those problems,

will not impinge on the responsibilities of the implementing ministry.

LOCATION OF THE MONITORING UNIT

It is for the above reasons that it is also desirable that the central unit be located in a position of influence within the administrative hierarchy. In the practice of the units surveyed to date, each unit was located in a powerful and influential agency within the administration. This location was in the Office of the Prime Minister or President (two), the Ministry of Finance (three), the Central Planning Agency (two), and a separate Ministry (one unit only, but whose Minister was also the President). In the case of both Central Planning Agency locations (Indonesia, India) the units were primarily gatherers and reporters of project performance information. Corrective action was the responsibility of other agencies in the central administrative structure.*

There is in fact, an overriding reason for locating the unit within that organization with the strongest influence on the development process. The political leader of this agency would presumably have the influence to support his monitoring staff in their analysis of implementation problems. Project implementation difficulties can arise due to the shortcomings in powerful ministries of government. A frequently recurring implementation difficulty, for instance, is the unplanned management of the Ministry of Finance's cash flows, a problem

---

* It might also be noted that neither of the Central Planning Agency units undertook any evaluation, although India did feed information on levels of achievement into the generation of the next year's annual plan.

that surfaces within projects in a multitude of different guises. Other centrally located problems arise in systems for customs clearances, in import licences, in the assignment of key personnel, etc. A monitoring unit that does not have the status and the political support to investigate these issues will accomplish little by reporting them (or their external manifestations) to the implementing ministry. The implementing ministry may be just as powerless to correct these problems as is a monitoring system that is used solely to identify them.

Another approach to the resolution of the more complex administrative issues is through a central management improvement unit. Most countries have such a group located in its central administration, most commonly within the Civil Service Commission. The influence of such units will depend on its staff capabilities and its political and administrative location, but typically this influence is not strong. Close cooperation with and support of this unit's activities by the monitoring and evaluation unit, however, will obviously strengthen a country's ability to identify and resolve major countrywide impediments to the implementation of its development activities.

The problems of a Third World country are problems of scarcity; scarcity of capital, administrative capability, materials, skilled labour, etc. Effective development requires that these scarce resources be planned and allocated to the various projects and programs throughout the country on a basis that is more centralized than in a country with a more abundant supply of resources. Problems that occur in this resource allocation and implementation process, therefore, may

possibly be solvable at the level of the implementing unit, but are much more likely to have arisen in the central processes of government. Attempts to resolve these problems therefore, are likely to be more effective if undertaken by a central unit rather than a decentralized one, be it a monitoring and evaluation unit with problem solving capability, or an administrative improvement unit within the central government.

THE DATA GATHERING AND REPORTING PROCESS (MONITORING)

There are a number of choices available to the units for gathering and reporting data during the implementation phase.

Management theory establishes three parameters to monitor in the execution of the project or program - the timing of the input supply, the cost of the inputs, and the performance level of inputs and outputs (quality, quantity and impact). In conjunction with the demand for the outputs, these factors determine whether the project will optimize its expected benefits and costs. But assessment of the impact of actual demand, and in many cases the impact of inputs and outputs, can usually only be carried out by an evaluation study. Routine monitoring therefore, usually covers only cost and schedule parameters, and in some cases, input and output performance.

Knowledge that the project, or parts of it, are behind schedule or otherwise not performing as expected, is not useful without two additional pieces of information: (i) knowledge of the problems that

are causing poor performance, and (ii) whether the poor performance will have a negative impact on the final project. The former information is necessary in order to help identify the actions that must be taken if the difficulties faced by the project are to be corrected. In many cases, the latter information can be incorporated into the routine monitoring system, although in practice few of the units did so.

The monitoring system therefore, would cover the timing and the cost of implementation, the problems that are being experienced, and to the extent possible and feasible, the performance levels of the inputs and outputs of the project. Details of the reporting processes that were used, together with other relevant factors in the data gathering process, are outlined below.

Periodicity of Monitoring

Six of the eight units examined to date operate on a quarterly reporting period, the seventh on a two-monthly basis, and the eighth operates on an 'as required' basis. This eighth unit, Peru, had developed an approach to monitoring and evaluation that differs from the other seven. It is an investigative unit within the Ministry of Finance which visits major projects within the country and, after a joint review of the performance of the project with project staff, issues a report to the Minister of Finance. After review and discussion in Cabinet, corrective action is in the hands of the responsible Minister. The unit also undertakes major review of administrative issues, such as Civil Service staffing levels, the viability of the larger public sector

undertakings, etc. The Central Planning Agency in this country also operates a routine monitoring system, but possibly because of the dominance of the Ministry of Finance, this system lacks impact. Because Peru does not operate a routine system, however, only seven countries have been reviewed in the examination below of the data gathering and reporting mechanisms that are used.

Schedule Information

All the monitoring systems measured actual completion dates against a pre-determined date, either for the completion of the whole project (usually only small projects), or for parts of it (measured against the completion dates of specific items of work, or against the number of physical units scheduled to be supplied or completed by the reporting period).

The method of establishing the dates varied. In most cases, the schedules were supplied directly to the monitoring unit by the implementing agency; in other cases they were the result of a joint planning activity. In the Honduras unit, whose routine monitoring covered only the physical components of projects and programs*, the planned schedules were drawn from comprehensive network schedules developed jointly by the staff of the monitoring unit and the projects. These schedules were monitored on site by staff members of the central unit attached to the project.

---

* Although the unit did review aspects of the administrative or economic impact of projects as a separate activity.

This unit's use of advanced scheduling techniques also enabled it to predict the impact of delays on completion dates. It may not be of importance to report items that are behind schedule depending on their impact on the final completion date of the project. Few units had the ability to determine this final impact.

Cost Data

Most units were unable to report cost data, primarily because of constraints on data collection. Generally, the time required to collect cost data was longer than reporting cut-off dates. As an essential feature of any monitoring activity is an early identification of the problems on which action needs to be taken, the value of including cost information was outweighed by the delays that would be caused by gathering and reporting it.

It was noted that the units that were able to report cost data were all small countries with a centralized disbursing system. Expenditure data, therefore, were centrally collected and readily available to the operating ministry or to the monitoring unit. Although a number of the units intend to introduce a cost control function in their system in the near future, in each case it entails a significant readjustment of the public accounting system.

The methods used to report costs were through a measurement of actual costs against the budget for the item of work, or against the budget for the quarter or for the year. If the actual work content is

comparable with the amount that has been used to build up the budget estimates, then such comparisons can provide an approximation of possible budget underruns or overruns. A refinement to this process, however, is for the monitoring report to include an estimation of the extent to which underruns or overruns incurred during implementation will impact on the annual or final project budget. Only one monitoring unit that was concerned with monitoring budget expenditures had this refinement.

Performance levels of inputs and outputs:

Although in no case did a central unit measure the quality of inputs or outputs as a routine process, nevertheless a problem reporting component does provide an opportunity to report deficiencies in the quality aspect of input or output performance. For instance, the report might identify poor quality seeds, diseases in the produce, inadequate road surfacing, poor motivation of staff, etc., as problems, all of which are quality parameters.

Quantities of inputs and outputs, when measured, were usually measured against a predetermined target or plan estimate. Outputs, when measured, were measured in only terms of those intermediate outputs (i.e. acres sown, families served, children taught, etc.) that were part of the program's efforts towards the achievement of a higher level objective or goal. As noted earlier, very few of the units had the evaluation capability to determine whether these higher level goals were

being achieved, or were likely to be achieved, or even whether the intermediate outputs being achieved were an optimum.

Problem Identification

Again, the approaches used to identify problems occurring during implementation varied considerably. The two units whose prime objective was monitoring broad plan implementation progress did not receive any formal reports on implementation problems. Of the five remaining systems that included a routine reporting component, all provided for problem identification. Three incorporated an unstructured system of notifying the monitoring unit of problems that were arising. In these cases, the project manager had space in his report to identify the problems that he was experiencing. The other two, both computerized systems, incorporated a highly formal process of problem identification. In one of these cases, the project manager identified the problem from a list of almost 200 coded problems. He also had the ability to identify the steps needed to correct his current difficulties. In the second case, a list of 15 coded problems was used. This latter system was, however, not too effective, as insufficient detail could be provided on identifying actual difficulties.

A coded problem represents a simple way of reporting project difficulties; it also minimizes some of the data processing effort. However, it does not always provide a clear identification of the problem, either because the project manager is not fully aware of the

reasons behind the problem, or if he is, a simple reporting of a coded problem provides insufficient detail. Reporting of problems by code number has the advantage of quickly identifying repeat problems, but such reports should also provide an ability to describe any issue in detail, including an ability to suggest (or request) corrective action.

The Analysis of the Data

The reasons behind the monitoring of the three parameters of cost, schedule and input/output performance are that these parameters are interchangeable, one with the other, in determining ways in which the project may be brought back to an optimum achievement of objectives. A project that is considerably above budget may be reduced in scope (output performance). Delays may also be compensated by a change in the timing, quantity, or quality of input supply (and, therefore in project cost and benefits). Schedule delays may also require a reduced output, or a slower than expected rate of achievement of benefits. A lower than planned achievement of project outputs may also be compensated for by a change in the timing, quality or quantity of inputs, again with consequences on the eventual benefits and costs of the project.

The types of analyses undertaken on monitoring data will depend on the type of project or program, and will often require the gathering and analysis of additional data. Such analyses entails the replanning of the input/output mix for the project, and would not normally be the responsibility of the monitoring and evaluation unit. In most cases, decisions on changes in the program were taken by the on-site project

manager, or by his parent agency. Two of the units included in the survey did undertake this work, although in conjunction with the implementing ministry.

The inclusion of a problem reporting component can simplify these analyses. The project manager may know the reasons behind poor performance, and may also know the steps that need to be taken to bring about improvement. Inclusion of this information in his report may facilitate a more rapid problem resolution.

Staff number in the monitoring unit

With two exceptions, all monitoring units operated with a staff of less than ten, and averaged about six professionals. The two exceptions were computerized systems, but Honduras, the smaller of the two, also provided staff to undertake direct planning and monitoring activities on the projects. This unit used about a dozen people for its central monitoring and evaluation activities, and about thirty staff located on the major projects throughout the country. The other unit, Malaysia, was a fully automated system covering 20,000 projects, employing about 100 professional staff.

The optimum staff level varies, of course, with the range of functions carried out and the number of projects and programs covered. In assessing staff requirements, it could be misleading to generalize, but it would seem that a dozen professional staff would be sufficient

for most monitoring unit to provide a range of services to the implementing ministries and to the central agencies of the government.

The Accuracy of the Data and the Response of Project Reports

All of the systems that monitored performance on a routine basis relied on the provision of performance data directly from the project. In a few countries, it was apparent that not all the project managers reported performance data that reflected adversely on their own performance, nor when they did, was it always accurate and reliable. Four factors appeared to influence the extent to which the project manager supplied prompt and reliable data. These were:

i) The status and position of the monitoring unit in the administrative hierarchy of the government.

ii) The extent to which the project manager believed that the system was not an audit of his own performance, but that it was designed as a system to help overcome some of the difficulties that he was facing.

iii) The extent to which he saw that monitoring actually assisted him in resolving issues. To this end, a number of the systems incorporated, either formally or informally, a problem solution status analysis. Through this system the project manager was made aware of the progress being made in resolving his difficulties.

iv) The quality of the cost and schedule control systems that were in use. On medium to large projects, the project manager's report of performance data would be more reliable if he used a modern schedule and cost control system.

Another method used to improve the quality of the monitoring data was to train the implementing staff in project management techniques. As has been seen, the systems of implementation planning and control that were used vary widely. In most countries, one of the central monitoring and evaluation unit's functions was to develop the planning and control systems that were used on different projects throughout the country, and to train project and ministry staff in their use.

IN CONCLUSION

Central monitoring and evaluation has the potential to assist materially in improving the rate and the effectiveness of development activities in Third World Countries. Despite the variation in techniques and methods used, many of the units that have been examined to date would appear to be making a contribution to this process. In some cases the contribution was positive and clearly observable. In other casas, the units seemed to be little more than data gatherers and reporters, with little measurable impact on implementation. The variety of systems used, and the variable impact that the units have, do make it difficult to assess, other than in subjective and impressionistic terms, which types of systems had the greatest impact on the rate of

development implementation. In this respect, it was clearly this writer's impression that units that became actively involved in the problem solving process were more effective. They were able to demonstrate more instances of problems that had been resolved, were in general more aware of the implementation issues faced by the projects in the field, as well as of those problems in the central government's administrative machinery that created difficulties in achieving effective planning and implementation.

A second impression, supported to a considerable extent by the reasons given earlier in this paper, was that a unit's effectiveness was directly related to its location in the administrative structure. Units which drew their support from strong positions of political influence seemed to be more effective than those without this support. Apart from these two factors, and the issue, also discussed earlier, of whether to cover a few projects in depth or to attempt to cover all projects, it would appear that the remaining choices in the operational methods of the unit (on systems of data gathering, analysis, reporting and problem solving) depend on the ways in which a government believes a monitoring and evaluation unit will best fit into its own administrative and political structure. These choices, however, will depend as much on the relationships among the ministers in Cabinet as it will on the efficiency of the alternative monitoring and evaluation systems.

Not all the units, however, are equally effective, and some of them do have weaknesses which reduce their effectiveness. For some of the units it is possible to delineate areas in which a further strengthening of their contribution would be possible. The principal areas of

strengthening would be in (i) a continuing elimination of any emphasis on a "policeman" or audit functioning of the unit, (ii) an upgrading of evaluation capabilities, particularly in impact evaluation but also in evaluation of process, and (iii) a strengthening of the management systems used nationally to control the planning, implementation and feedback on projects and programs.

A central monitoring and evaluation unit has the objective of improving the effectiveness of developmental efforts. Again, it is difficult to assess how useful the units are in comparison to other approaches to achieving that improvement. The problems of development implementation can be identified and resolved by other agencies and other systems. As mentioned, most governments have administrative reform units. It is also possible to strengthen the monitoring and evaluation activities within the line ministries. However, it is clear that ministries are part of a complex governmental administrative machinery. For the reasons outlined earlier, adequate monitoring and evaluation of the problems of implementation from a position outside this central administration is not likely to be successful. A central monitoring and evaluation unit is part of this central administration. With an appropriate organizational location and operating systems, it can provide the impetus and the methodology to solve some of the difficult problems that impede Third World countries in their attempts to implement viable and effective developmental projects and programs.

THE MONITORING AND EVALUATION FUNCTION

There are many definitions of monitoring and evaluation, ranging from one extreme which defines the entire process of monitoring progress and evaluating project impact as monitoring, to the other extreme that recognizes few management control activities other than the techniques of evaluation research. In this latter case, monitoring is defined simply as an information gathering activity.

Any analysis of monitoring and evaluation functions requires that the terms be clear and well understood. The definitions which are given below recognize a close interrelationship between monitoring and evaluation, and attempt to develop a definition based on the methodology which, in practice, would appear to distinguish the two functions.

Monitoring: A routine gathering and reporting of performance data (both input and output) against measures of expected performance, the analysis of which may lead to the taking of the administrative action necessary to replan the activity, or reorient it towards the achievement of its origional objectives. Monitoring can be concerned with the provision of the physical infrastructure, or with the operation of the project or program.

Evaluation: A discrete study (although the study may subsequently be repeated), or analysis of project or program data designed to determine if the inputs are optimizing the

achievement of objectives. The study may suggest corrective decisions for the planning and implementing of the current or future projects. Evaluation arises during the monitoring process (on-going evaluation[1]), or it may be a separate activity in its own right (ex-post evaluation[2]).

Some examples: A controlled research survey of the knowledge, attitudes and practices of a group of farmers in an area that is, or has been, the target of particular rural developmental efforts is an evaluation study. A study of the project's impact on the incomes of these farmers is also evaluation. The attendance by the project manager at a meeting with the farmers in which he tries to determine how the farmers are responding to the program, and whether he needs to make any changes in it, is part of his normal management function, but it is also a monitoring and evaluation activity.

The checking of progress towards achievement of the objectives of a major new fertilizer project, is monitoring. The gathering of monitoring data, however, is of no purpose unless it leads to the management action necessary to correct any problems that have been

---

1 Evaluation of the impact of a project or program (or aspects of it) that is still underway.

2 Evaluation of a project that has been completed. There is a third type of evaluation, ex-ante evaluation, which is concerned with the analysis of the project during the planning stages. Other than through the use of ex-post evaluation data from earlier studies, this type of evaluation is not connected with the evaluation processes described in this document.

identified. The decision on the best action to take may, of course, require additional data and additional analysis. In this paper, if the analysis of monitoring data and the taking of corrective decisions is part of the routine management of the project, the whole process has been considered as monitoring.

Any significant analysis, however, of the need for major changes or redesign during construction of the project (including changes in layout, equipment, etc., that could lead to a revised economic objective) would be evaluation. For an on-going program or project, the routine gathering of data on yields (from a rural development program, for instance) leading to the taking of any corrective management decisions is monitoring. Any major analysis of the significance of this data, possibly leading to recommended changes in the program, however, would be evaluation.

In other words, monitoring and evaluation are close, even partially overlapping management functions. In addition, there is little conceptual difference between a project manager stopping to analyse the actual or potential achievements of his program or project, and an elaborate research-based evaluation study. The difference is only in the techniques used. Both are also management tasks and both have the same objective - to check if the mix of inputs to the project, or the project objective, needs any adjustment.

Evaluation (as a separate exercise) can also be divided into two categories: (i) administrative or process evaluation and (ii) socio-economic, or impact evaluation. Once again the line between these two categories is blurred. Administrative evaluation examines the effectiveness of the project's organizational and managerial systems, and of the methods and processes used to implement the project. It also develops proposed changes which would enable the project to optimize the achievement of its objectives. Social and economic impact evaluation examines causal relations between the mix of inputs to the program and its outputs. Impact evaluation determines the extent to which benefits, in relation to costs, are achieved, but in this case, benefits primarily derived through the achievement of social or economic objectives, rather than through an effective implementation process. Once again it will develop recommendations on changes that will enable the program to meet existing or revised objectives.

An evaluation study, therefore, may have administrative, social, economic and technical implications. For example, an examination of the distribution system for contraceptives may question the impact of the method of distribution on the acceptance levels of actual or potential users (social), the implications of different storage methods and physical distribution facilities (administrative, technical), the effectiveness of the contraceptives themselves (technical) and their effectiveness, along with other factors, in preventing births (economic).

There are two reasons for making the above distinctions in the types of evaluation that are carried out. One is that in many of the units studied, there was a tendency to evaluate administrative aspects, but rarely social and economic implications. The distinction, therefore, provides a ready method by which the evaluation practices of the units can be categorized. The second reason is that different skills are required to evaluate administrative implications, as opposed to social or economic implications, a difference which has significance for the precise functions and location of the units, as is discussed in the text.

VIII A REVIEW OF EX-POST AND EX-ANTE EVALUATION ANALYSES

This chapter goes beyond the scope of the workshop and is included for future reference.

It consists of exerpts from a report originally prepared by G. Edward Schuh and Helio Tollini* for the Consultative Group for International Research. The first exerpt discusses some definition and measurement problems particular to agricultural research.

SOME CONUMDRUMS

Difficulties discussed in this section are of a more operational nature. They involve both problems of concept - how one defines what one is after - and problems of measurement. Each presents difficulties in making cost-benefit analysis an operational tool for guiding the allocation of agricultural research resources.

a. Definition of output

In principle the research process can be viewed just as any other production process. Inputs of various kinds are combined in rather particular ways to produce an output. The problem is in knowing how to define the output.

At the most abstract level, most would probably agree that the important output of the research process is new knowledge. If there were a market in which this new knowledge were bought and sold, and the output were in identifiable units, the measurement problem would be relatively straight-forward. One could measure the number of units produced and multiply it by the price determined in the market, and a measure of the total value of output would be at hand.

---

*Costs and Benefits of Agricultural Research: State of the Art, and Implications for the CGIAR, 1978. Also World Bank Working Paper 360, 1979.

Unfortunately, the world is not that simple. Knowledge is intangible and in some respects undefinable, although in many respects we may recognize it when we "see" it. But reaching agreement on what a unit of knowledge is would be quite difficult.

Similarly, no well-defined market exists for knowledge, despite the fact that knowledge obviously has economic value, and that individuals who produce it tend to be rewarded in proportion to their ability to produce it. Hence, at this rather abstract level we are left in a situation in which it is difficult to agree on the unit we want to measure, and there is no mechanism which gives us a direct measure of the value of the unit to society.

Because of these difficulties, researchers attempting to evaluate research have used a number of different surrogates for the output of the research process, each of which has certain strengths and weaknesses. One approach is to recognize that the knowledge produced by the research process is typically published in the form of scientific and technical papers. The number of such papers then serves as a proxy for the output variable. 1/

An advantage of this approach is that it does provide a number to work with. Moreover, to the extent the publications are screened by knowledgeable scientists as a basis for publication, there is some assurance that a publication reflects something called "new knowledge".

By the same token, however, it is clear that the "amount" of knowledge transmitted in a published paper varies a great deal from one paper to another. There are serious difficulties in attempting to place relative values on the publications, and the most that has been done to date is to assume that "a publication is equal to a publication".

It should also be noted that the review process for publication is itself imperfect. Well-intentioned researchers can disagree over whether a given publication contains information that advances our knowledge or not. We are all aware of the unorthodox but important idea that is rejected for publication. And similarly, we are all aware of the paper that gets published because of the name attached to it, rather than the content it carries.

The use of number of publications as a measure of the research output can also have pernicious long-run consequences on the research industry, even though it may have value in an ex post check of what has transpired. The problem is that researchers can easily play the "publication game" if numbers of publications are used as an index of output and productivity. Publications can be broken down into smaller units, different perspectives on the same project can be published in separate journals, and so on.

1/ See Chapter 5 of Evenson and Kislev.

An alternative approach is to define the output in terms of some well-specified innovation. This moves the concept of output closer to something that has more immediate economic value, and it provides an index about which it might be easier to reach agreement. Moreover, an innovation can be related more directly to the original objective or objectives of the research process. Examples of such innovations include hybrid corn, an improved inbred line, an insecticide that treats a particular pest, a fungicide that controls a disease, etc.

The difficulty with this concept, of course, is the problem of non-adoption. In some sense an innovation has to meet a market test. If the innovation is not adopted because it is too expensive, as for example in the case of a pesticide, the research has not attained its full objectives, even though in terms of a particular research project it may have attained its goal. In this sense, to measure the output of the research process in terms of innovations generated can be misleading, or at best not tell all of the story.

For this reason, most attempts at evaluating the output of a research process have concentrated on measuring that output in terms of its impact on the production process. As will be noted below, this involves making some independent estimate of the extent to which the research has shifted the production function or the supply function for the particular crop or livestock category. The analysis is formulated in terms of a particular innovation, but the empirical work is in terms of resources saved, shifts in the production function, or shifts in the supply function. The economic calculus can then be applied to estimate the flow of benefits.

A particular example of this approach is the evaluation of the social rate of return to agricultural research by means of the Hayami-Ruttan metaproduction function. 1/ This production function includes social or public inputs in addition to those used directly on the farm. As a measure of the research input, the number of scientists working at agricultural research, or at research and extension, is introduced as a separate variable in the function. 2/ By statistical means, then, the contribution to agricultural output of this particular set of inputs can be isolated, and an estimate of either the productivity of the research inputs or the social rate of return to investments in research can be made. Similar to the case immediately above, the output of the research effort is measured in terms of its <u>effect</u> on agricultural output.

Despite the complexity which has emerged from the above discussion, only the simple case has been treated - the case in which the principal contribution of the research is expected to be an increase in agricultural output. More complex outputs of agricultural research include qualitative changes in the product, such as a tomato that transports better, or an improvement in the nutritional value of a product.

1/ For an example see Thompson (98).

2/ Alternative variables such as total expenditures on research can also be used.

Similarly, very little progress has been made to date in evaluating the returns to economic research. Here the gamut is quite wide: farm management research which leads to a more rapid adoption of technical innovations, marketing research which improves price incentives to farmers, research which leads to improved understanding of economic behavior, and policy research that leads to an increase in agricultural output from a given bundle of inputs due to the removal of policy distortions. Research to devise a suitable methodology for dealing with this class of problems has not made much progress, largely because of the difficulties encountered in defining the output of the research process.

Finally, there is the problem of maintenance research. Certain agricultural innovations are subject to a relatively high rate of obsolescence. A new wheat variety becomes susceptible to a new strain of rust. Insects develop resistance to insecticides. And antibiotics lose their effectiveness in combatting a particular disease.

The consequence of this obsolescence is that a certain amount of ongoing research is required to just "stay even," in contrast to advancing the frontiers of knowledge or even the efficiency frontier. The output of such maintenance research is just as important as research which pushes the knowledge frontier further out. 1/ Identifying it and taking it into account may be somewhat more difficult, however.

b. Definition of inputs

For certain categories of inputs the definitional and measurement problems on the input side are more straight-forward. For example, the research process typically involves the use of certain labor inputs, including skilled professionals, physical inputs such as land, buildings, and capital inputs, and agricultural inputs such as research animals and fertilizer. With appropriate accounting procedures these inputs can be measured with a fair degree of accuracy, although difficulties arise when it is necessary to attribute fractions of the services of skilled labor and capital equipment to particular research endeavors.

The difficulties arise in knowing how to conceptualize and measure the inputs of certain skilled manpower, and in knowing how to treat the existing stock of knowledge. In treating skilled manpower, two problems arise. If labor markets were efficient and without distortion, the salary paid the scientist would be a suitable measure of the value of his services. However, labor markets for skilled manpower are neither efficient nor free of distortions. Hence, in some cases a more suitable approach would be to attempt to shadow price the labor input. 2/

1/ Ayer and Schuh (10) argue that the high payoff to the cotton research program in Brazil was due in part to the speed with which it developed varieties that were resistant to a devastating infestation of wilt.

2/ Shadow prices are measured either implicitly, or in terms of the opportunity cost of the input. They are used when observed market prices do not reflect the true scarcity value of the resource to the economy.

The second problem is to know how to measure the chance idea that comes from an encounter with another researcher. To the extent that this is a free good, there is no problem. It enters as a free good. To the extent a consulting service is involved, the question becomes more complicated if the labor market is not efficient. These problems in pricing the services of scientific manpower are for the most part at least tractable. The more difficult problem is to know how to price and measure previous research. The very successful cotton research program in Brazil had its start with lines brought from the U.S. 1/ To what extent should the cost of producing those lines be considered in evaluating the benefit-cost ratio in Brazil? The procedure followed by Ayer and Schuh was to treat the lines as a free good.

Similarly, there was considerably research in the U.S. and Japan that was drawn on in producing the Mexican wheats. How far back should the analyst go in imputing the costs of these programs to the Mexican wheat program? If the outputs of these previous research endeavors are treated as a free good, the social rate of return to the Mexican wheat program would be exceedingly high. If all the costs leading up to that particular innovation were included, the return would be substantially less.

When the research endeavor is well-defined and institutionally specific, as in the case of the Mexican wheats, operationally the answer seems relatively straight-forward. But if the Mexican wheats had by chance been produced as a logical consequence of either the U.S. or Japanese endeavors, the answer would not have been so clear. This brings out the difficulty in knowing how to treat the important input of past knowledge. 2/

A similar although somewhat different problem arises when one considers a research organization like the land grant colleges. Typically, a considerable amount of basic research will be going on in one part of the university, while the school of agriculture will be concentrating more heavily on the applied side. The synergism among researchers can produce a substantial amount of knowledge that is "free" to the applied research program. The question is how to take account of this when assessing benefit-cost ratios for the research program. If only the inputs involved in the applied program are considered, the results will be of one dimension. If all research at the university is included, the answer will be quite another. To date there are no easy answers to this important imputation problem.

Evenson, Flores, and Hayami (33) have handled this "transfer" problem by regressing the change in yield on a series of knowledge stock variables that represent research activities in related disciplinary

---

1/ See Ayer and Schuh

2/ It should also be noted in this context that the output of the Brazilian and Mexican programs cited are logically outputs of their predecessor programs. The conceptual and measurement problem in some sense, then, is a problem of identifying and measuring the value added from the respective programs.

and commodity programs and research activities in other geographic areas that might be transferable. Hence, statistical procedures are used to isolate the separate effects of the various programs, and the costs and benefits can be computed accordingly.

A final problem with inputs arises outside the research process per se. The adoption of a new innovation can induce shifts in resources among sectors. If one is using secondary data to estimate the benefits of the research program, some attempt has to be made to take account of these resource shifts. A difficulty arises because labor is often an important component of the resource shift, and labor markets often are imperfect. Consequently, the observed wage rate (or rates) may not be an accurate measure of the true opportunity or social costs of the labor. Some method of shadow pricing then has to be used.

c. Tangible Output of the Research Process Versus Effective Adoption at the Farm Level.

This problem was discussed briefly above. But a somewhat more systematic treatment might be helpful. One way to assess the cost-effectiveness of a research program is in terms of the objectives of the research effort. These could be specified in quite precise and easily understood terms: (1) to produce a variety or varieties of a crop that has double the yield potential of existing varieties, (2) to find a means of controlling a particular insect, or (3) to find a means of controlling a particular disease. A management-by-objectives approach to the assessment and monitoring problem would then assess whatever cost it took to attain the specific objective.

But the mere attainment of the objective - which typically can be expressed in terms of an innovation at the farm level - does not guarantee that attaining that objective is relevant to the farmer. For example, the crop variety with double the yield potential may not fit into existing cropping patterns, or as with the early IRRI rice varieties, the new plant may not have suitable consumer qualities. In the case of the high lysine gene in corn, the research objective of introducing the gene was attained. But high lysine corn is still not a commercially viable enterprise because yield objectives have not yet been attained and there are problems of product identification or differentiation in the market.

Similarly, the system for controlling an insect may be too costly in terms of chemical and labor inputs, or it may have undesirable environmental consequences. The same applies to new means for controlling a disease.

One solution to this problem, of course, is to place the appropriate specificity on the research objectives. Hence, the goal would be to attain economically viable varieties and systems of insect and disease control. Similarly, the appropriate environmental constraints can be placed on the research objectives. Although feasible, the general use of such specificity would for the most part be a departure from present practice.

In the final analysis, innovations have to meet the market test when an attempt is made to evaluate the cost effectiveness of a research activity. Equally as important, they have to be consistent with whatever institutional restrictions the body politic has put on innovative activities. These considerations suggest that the cost-benefit or cost effectiveness analysis has to be made in economic terms and with institutional constraints, in contrast to the more common operational objectives of a research project.

d. Ex ante Versus Ex post Considerations

Benefit-cost analyses of agricultural research can be made either in an ex post sense or in an ex ante sense. To date, most of them have been made from an ex post perspective, since their primary objective has been to assess the role of agricultural research in economic development, and to determine whether investments in such activities have been economically viable. We are aware of four ex ante analyses of agricultural research, 1/ although as will be noted below there is a stronger tradition of ex ante analyses for the industrial sector.

Which of these is desired depends on the purpose of the assessment or monitoring. If the goal is to determine how efficient particular research institutions are, or how efficient particular lines of research activity have been, then an ex post approach is required. If the goal is to use benefit-cost analysis as a guide to how research resources should be allocated to maximize their payoff, an ex ante perspective is required.

Although not used on an extensive scale for this purpose, we believe substantial emphasis should be given to ex ante analyses. After all, an important question is to know how many resources could be allocated to attain a given research objective. Moreover, one would like to know before the fact where the expected payoffs are likely to be the greatest. This, of course, is not to deny the value of ex post analyses, for they can provide important insights into the research process, and provide a basis for comparing alternative organizational and methodological approaches.

Somewhat different methodologies are required to make an ex ante analysis than is required for an ex post analysis. Moreover, the preciseness and robustness with which inferences can be drawn are quite different in the two cases. We will return to these problems below.

e. The Role of Economic Policy

The role of economic policy has not been given a great deal of attention in past benefit-cost analyses of agricultural research. Ayer and Schuh (10) and Schuh (83) have noted the role of economic policy in determining the distribution of benefits from research between consumers and producers in the society. But only Hertford (49) and his colleagues have to

---

1/ Castro and Schuh (75), Easter and Norton (29), Klein (53), and Pinstrup-Andersen and Franklin (71).

our knowledge systematically addressed the question of whether economic policy affected the rate of return or the benefit-cost ratio of research.

Clearly, economic policy can affect the measured benefit-cost ratio of research, and in a number of important ways. First, to the extent that economic policy distorts price relatives it may cause a failure to adopt innovations that would otherwise be adopted. An important example occurred in Brazil when trade and other policies lowered the domestic price of agricultural products relative to the price of fertilizer. Since under the circumstances the use of fertilizer was not profitable for many crops, a considerable amount of soil and fertilizer research went for nought.

The consequences did not stop there, however. Since the use of fertilizer on corn was not profitable under the prevailing price relatives, farmers also failed to adopt hybrid corn. Hence, the return to what was a relatively effective research program on hybrid corn was quite low.

Clearly, economic policy caused Brazil to undervalue its research activities. What might have been a relatively viable research effort was perceived as a low payoff activity. Hence, Brazil invested less in agricultural research than it might have under a different policy regime, and sacrificed this important source of growth.

The role of economic policy in influencing the observed productivity or return from agricultural research can perhaps be seen more directly in a somewhat different context. If the goal of a research program is to increase the output of a particular crop or crops, the benefit-cost analysis or evaluation of cost effectiveness may require placing a value on the increase in output. If economic policy has distorted relative price ratios, the contribution of the research program can be either under- or over-estimated, depending on the direction of the distortion. 1/ To evaluate the benefit-cost ratios in this case some estimate of shadow prices is required.

Economic policy has still other effects on the perceived or actual benefit ratios of agricultural research. For example, policy measures which restrict exports of agricultural exports can reduce the earnings potential of the research in terms of exchange earnings, thereby causing the actual benefits from the research program to be substantially reduced. The reduction in foreign markets can reduce the income and employment generating potential of particular innovations. And trade policy can influence whether it is low-income consumer groups that receive the benefits of research or upper-income producer groups. Each of these are factors that should be taken into consideration in assessing the productivity and cost-effectiveness of research. Clearly, the particular economic policy regime has a major impact on the potential, perceived, and actual returns from agricultural research.

1/ India has substantially over-valued wheat in the domestic economy, compared to international opportunity costs, while substantially undervaluing rice. This has undoubtedly affected the perceived rates of return to its respective research programs, and probably has affected the relative rates at which new innovations have been adopted. See Sukhatme (95).

f. Negative Results

The effectiveness of agricultural research is typically evaluated in terms of its successful contribution to specified research goals and objectives. An important conundrum is how to handle the problem of negative results. Several years of concentrated effort may have failed to increase the yield of a particular crop. But the research effort itself may have clearly shown that a number of possible approaches to raising yields is not viable. The problem is to know how to handle such negative results.

Clearly, there has been an increase in knowledge, for researchers now know at least part of what will not work. This is of value to present and future researchers. But in terms of atteining the operational goal of the research, it has not made a contribution. Ultimately, of course, this problem comes down to the concept of output used to measure the effectiveness of the research program. In terms of additions to the stock of knowledge, the research effort may have been quite productive. In terms of an on-the-farm gain in agricultural output, little may have been accomplished.

g. Joint Outputs

The final conundrum has to do with the fact that a research program inevitably involves joint outputs. For example, researchers acquire new skills and knowledge by the very process of engaging in research. This increase in human capital has considerable value to society and should be taken into account in assessing the productivity of the research effort.

Similarly, research programs typically have a training or educational program associated with them. The complementarity between teaching and research is well recognized, and the teaching is assumed to be better if it is associated with a vital research program. Some part of the educational output should therefore be attributed to the research program.

Similarly, there are important institutional spillovers from an effective research program. The contribution of the International Centers in strengthening national research programs and systems is an important example. The knowledge, skills, and experience imbedded in the researchers at the Centers presumably have value to the national programs. Similarly, a researcher may take on special value as a consultant to government or other researchers as a result of his participation in an effective research program. These contributions can be an important output of a research program, and should be taken into account in assessing its productivity. To date, however, such considerations have not been included when evaluating the effectiveness of research programs.

This second exerpt from the Schuh and Tollini report reviews methods and procedures used in evaluating the effectiveness of agricultural research. These methods go well beyond ascertaining whether a specific program is being implemented as planned or has achieved its expected output. They assess either the contribution that program(s) have made to reaching overall development goals, or the potential contribution that proposed program(s) could make.

METHODS AND PROCEDURES FOR EVALUATING AGRICULTURAL RESEARCH

This section contains a review of various methods and procedures that might be used to evaluate the effectiveness of agricultural research programs. The material is organized under two headings: (1) procedures used to make an ex post evaluation of the effectiveness of agricultural research, and to evaluate the contribution that agricultural research has made to the economy as a whole, and (2) procedures that have been used to make an ex ante evaluation of research proposals and programs, largely with the goal of developing a more effective means of establishing priorities but also for purposes of justifying budget requests. The examples cited range from broad sectoral studies to narrow evaluations of particular research projects.

The purpose of the discussion is not to make a critique of past procedures, or to enter into a discussion of the intellectual niceties of particular procedures. Rather, the objective is to provide the reader with a general notion of the procedures that have been used in past studies, the data required to implement them, and some of the issues involved.

A. Ex Post Studies

Studies that have attempted to make ex post evaluations of agricultural research have for the most part focused on output-increasing technologies, and have neglected other contributions of agricultural research. The procedures used can be grouped into five different classes: (1) those which attempt to estimate the resources saved by the adoption of new technology; (2) those which use the concepts of producer and consumer surplus; (3) those which introduce investments in research into an aggregate production function; (4) those which estimate the impact of the technology on national income; and (5) those which have attempted to identify the effect of increased output on nutritional status of the population. We discuss each of these in turn.

a. The Inputs-Saved Approach

Professor Schultz (85) used this approach in what was to the best of our knowledge the first major attempt at quantifying the returns to investment in agricultural research. His interest was in US agriculture as a whole, and hence he made no attempt to consider individual research programs or particular technological innovations.

If this approach is used to measure resource savings over an extended period of time, an index number problem arises in measuring the value of inputs saved since relative factor prices change over time. To deal with this problem Schultz developed upper and lower limits for the resources saved by in one case using price weights from the early part of the period and in another case the price weights from the end of the period. The resource savings are then estimated by determining how many resources would have been used to produce the output of a base period using the techniques of production of an earlier period. A comparison of this with the resources actually used provides an estimate of the resources saved.

From this perspective, the value of the resources saved constitutes the benefits from the research. The costs of producing these benefits are then estimated by calculating the cost of all research and extension in the country—both public and private. A benefit-cost ratio can then be calculated, or the data can be used to make an estimate of the social rate of return.

This approach could be extended to individual commodity programs or to more narrowly defined technological innovations. Since aggregate data are not likely to be available in sufficient detail for such an analysis, however, it would be necessary to estimate resource savings either from experimental data or surveys of farms. When combined with data on the extent of use of the innovation, an estimate of the total resources saved could be made. Such an approach would be especially useful for evaluating innovations that are more directly resource-saving than output-increasing.

b. The Use of Consumer and Producer Surplus

An alternative approach is to measure the benefits and losses from technical change by means of its impact on what economists refer to as consumer and producer surplus. 1/ The technological innovation is assumed to shift the supply curve for the product to the right. When this occurs, consumers benefit from having more of the product available, and producers may benefit from the reduction in costs of production. The concepts of consumer and producer surplus can be used to measure these benefits (and possible losses).

The advantage of this approach is that it is relatively flexible, and in the hands of a skilled analyst can be modified to take account of a number of side effects of technological change, as well as some indirect effects such as the impact on trade, and the impact of trade and price policy on the distribution of benefits from technical change. The methodology is feasible, however, only when a set of improved technologies can be associated with a particular research program or programs. If there should be significant technological transfer among countries or regions, or from other research programs, it will not be possible to associate shifts in the supply function with a specific research program.

The basic analytical framework is illustrated in Figure 1. A shift in the aggregate supply curve (from S to S′) for the commodity under consideration is depicted, with this shift assumed to be attributed to improved technology. The shift in the supply curve produces a change in the consumers surplus by the area P_oABP_1 (the area ADB plus the area P_oADP_1). This is a flow of benefits that arises because consumers are able to purchase more of the product and at a lower price (in the general case). The same shift in supply will produce a change in producers surplus by the area BDO minus the area P_oADP_1. The total change in economic surplus (producers plus consumers) will be the area of AOB.

The simple framework provides the basis for analyzing various aspects of agricultural research. The empirical information required is knowledge of how much the technical change shifted the supply curve, and knowledge of the parameters that describe the conditions of supply and demand for the product. Information on the costs of the research program required to induce the shift in the supply curve is also required if a full benefit-cost analysis is desired.

1/ For an easily understood explanation of economic surpluses and their use in measuring the returns to agricultural research, see Hertford and Schmitz (48).

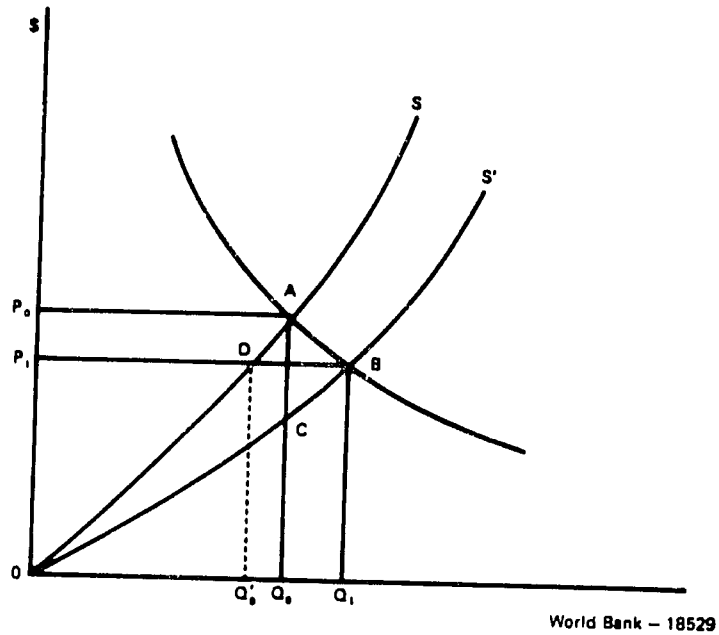

FIGURE 1
BASIC MODEL FOR THE ANALYSIS OF
CONSUMER AND PRODUCER SURPLUS

One of the most important advantages of this approach is that it provides a means of analyzing how the benefits of the research are divided between consumers and producers. The only thing required for such an analysis is knowledge of the demand and supply curves. As is readily apparent from the figure, producers can sustain losses from the technical change. All this requires is that the area P_0ADP_1 be greater than the area BDO. This distributive aspect is more important if policy makers should have as a particular goal an improvement in the welfare of either producers or consumers.

The model may be applied to a closed economy or to an economy open to trade. The demand elasticities in an open economy will, of course, tend to be quite high, meaning that there will be fairly small changes in price associated with changes in the quantity supplied. Hence, if the technical change occurs for a product that is being exported, most of the direct benefits of the research will go to producers, unless there should be government intervention. 1/ Consumers will benefit indirectly, however, since the added foreign exchange which increased exports earn will lower the price of imports and help to finance a higher rate of growth. These indirect effects should be taken account of in calculating the benefits and costs of research.

The model can also be modified to take into account price and trade policies, and their impact on the distribution of benefits from technical change. Schuh (84), for example, has argued that the over-valued dollar in the post-World War II period caused a larger share of the benefits of technical change in US agriculture to be transferred to the US consumer than would have been the case if exchange markets had been free. Akino and Hayami (3) have examined the rice breeding program in Japan, which was interesting in light of the fact that Japan was a net importer of rice during the period covered by their study. These authors give explicit attention to the distribution of benefits between consumers and producers, and conclude that in the absence of trade Japanese producers would have been net losers from agricultural research.

Evenson, et al. (33), consider the distribution of benefits from rice research in the Philippines. In that case, imports have been utilized to maintain a stable price for consumers, with sufficient rice imported to maintain a target domestic price. Suppose P_1 in Figure 1 is the target price. With the original domestic supply function, the quantity DB would have been imported. The shift of the supply function to S′ would eliminate rice imports. The consumers surplus for rice would remain unchanged, but producers would gain by the area ODB. This area represents a welfare gain to society and is equal to the change in the resources devoted to domestic rice production, $OBQ_1 - ODQ_0'$, plus the value of the imports in the initial situation, $Q_0'DBQ_1$.

These examples show the potential richness of the analytical framework. To the extent policy makers are interested in the effects of economic policies, in the distribution of benefits and costs from policies and

---

1/ For a more detailed analysis, see Castro and Schuh (75).

technical change, and in trade implications, the model provides a means of giving them some important answers. Moreover, these answers are quite important in establishing research priorities, and in managing the process of technical change for the greatest good.

The model can also be modified in a number of important ways to accomodate different problem situations. For example, Griliches (42), in his first use of the economic surplus concepts to estimate the benefits of technical change, assumed that all the benefits of agricultural research were realized in the form of a consumer surplus. Two alternative estimates of this surplus can then be made, depending on whether the supply curve is assumed to be perfectly inelastic or perfectly elastic (Figure 2).

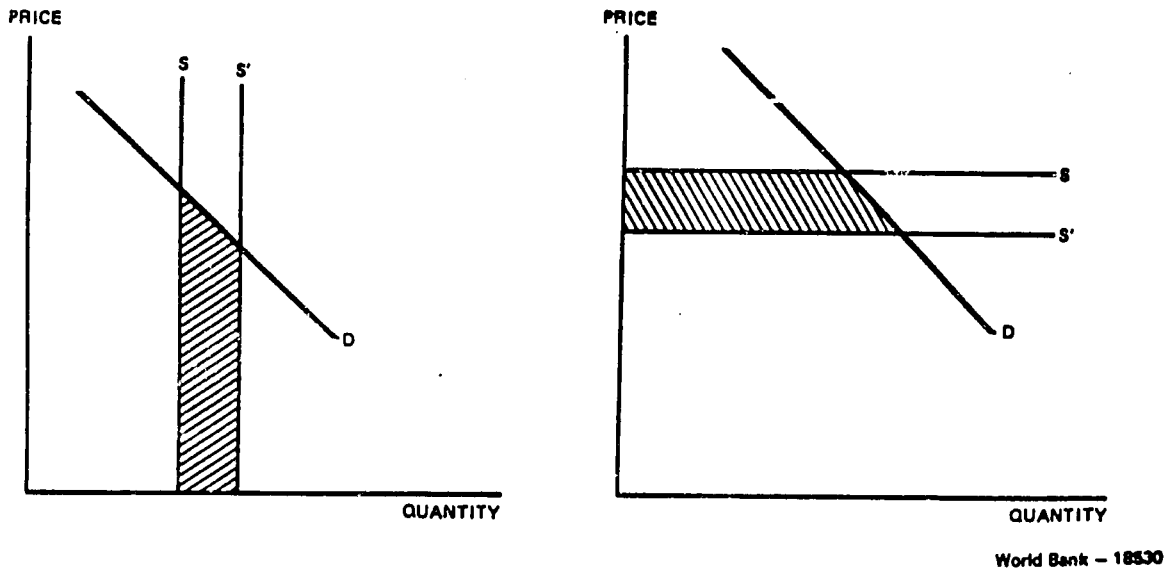

FIGURE 2

MEASURES OF CONSUMER SURPLUS UNDER ALTERNATIVE ASSUMPTIONS ABOUT ELASTICITY OF SUPPLY

The shaded areas define the alternative measures of the consumer surplus and are assumed to represent the benefits of the technical change that resulted from the research effort. Griliches argued that these alternative measures constituted upper and lower limits of the benefits of the research, although Lindner and Jarrett (57) have recently taken issue with that conclusion. Griliches also assumed a closed economy and ignored the potential foreign exchange that might be earned from technical progress.

The strength of the Griliches approach is its simplicity. It assumes that the price elasticity of demand is -1, and thereby abstracts from general equilibirum or resource adjustment problems. No estimates of either the demand or supply parameters are required, since they are all handled by assumption, and the trade sector and distribution consequences are also ignored. The major empirical problem is to obtain a measure of productivity gain that reflects only the output of research. This measure is required, of course, to know how much to shift the supply curve. 1/

Peterson (69) has shown how general equilibrium effects can be taken into account in estimating the net benefits of research. These effects arise by virtue of the changes in resource productivity associated with technical change, which in turn cause resources to be either induced into the progressive sector, or expelled from it. Peterson's procedure is based on the simple fact that if the price elasticity of demand were equal to -1, the total value of any price-quantity combination along the curve is the same. The relation between the unit elastic demand curve and the actual demand curve provides the means for taking account of the general equilibrium effects.

Consider Figure 3. Assuming that we shift the supply curve back to the left to make the evaluation, the decline in output q"q′ is due to a decline in productivity of a fixed bundle of resources if the technology were withdrawn. As a consequence of this decline, net social benefits decline by q"q′BB′. (These net social benefits are made up of both producer and consumer surpluses.) The decline in output represented by $q_o q''$ is due to a shift of resources out of the sector. These have an opportunity cost (as represented by the area under the supply curve) of $q_o q''G'G$. But there is also a loss in consumer surplus in the amount of $q_o q''B'G$. Therefore, the net gain due to the liberation of resources is G B′G′. This amount has to be subtracted from the estimate of net social benefits measured in the partial equilibrium framework to take account of the general equilibrium effect.

In the particular case considered, the actual demand curve was assumed to have an elasticity greater than 1 in the relevant range. If it had an elasticity less than 1, the relationship between the two demand curves would be different, and the triangle would represent a net loss, and have to be added to the partial equilibrium estimate of benefits.

1/ Benefits are always measured by assuming the production technology were withdrawn (a shift to the left of the supply curve) in order to provide a conservative estimate of the benefits.

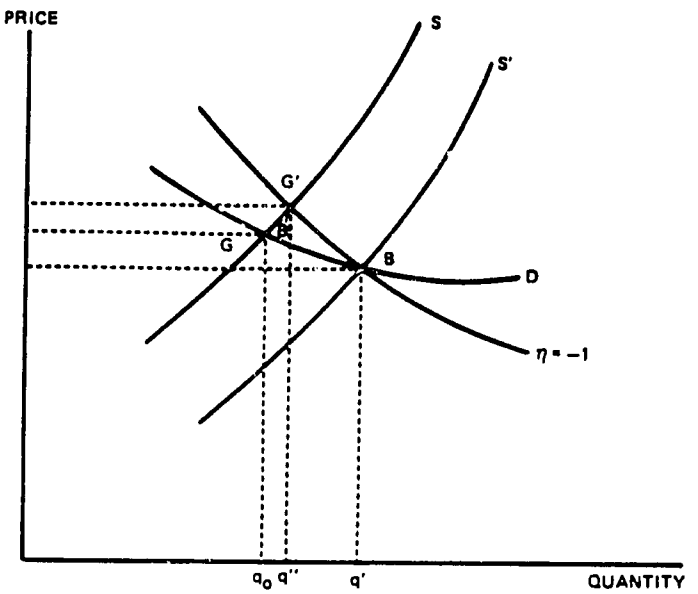

FIGURE 3

ILLUSTRATION OF GENERAL EQUILIBRIUM EFFECTS ASSOCIATED WITH SUPPLY SHIFT

World Bank – 19531

The Paterson formulation of the problem is somewhat demanding in terms of the measures of economic parameters required to make the estimate. By the same token, however, a great deal more can be learned from the analysis in terms of the distribution of benefits between consumers and producers and the adjustments for general equilibrium effects.

Schmitz and Seckler (82) extended this basic model in an interesting way. As noted above, the increase in productivity due to technical change can require the release of resources from the sector, depending on the conditions of demand for the product. If alternative employment possibilities are not available, the resources released may be unemployed. If they are, then the income lost by these unemployed resources has to be deducted from the benefits of the technical change to determine the net benefits.

Schmitz and Seckler viewed the mechanical tomato harvester as an innovation of agricultural research which caused farm workers to be unemployed. They first estimated the benefits of the research on the harvester in the usual way, and then estimated the returns the unemployed workers would have received in the absence of the technical change. These sacrificed returns were then subtracted from the benefits.

This procedure amounts to taking into account the adjustment cost associated with the technical change. It should be noted, however, that resources could be induced <u>into</u> the sector as a result of technical change. If this occurs, it should be taken into account as well, since it will mean a reduction in output in other sectors.

Ayer and Schuh (10) further extended the model by specifying a cobweb behavioral model for the sector. This model leads to a different kind of adjustment costs because <u>ex post</u> results do not square with <u>ex ante</u> expectations. Their procedures net out these adjustment costs from the net flow of benefits.

Ayer and Schuh also made a qualitative analysis of which group of factor owners received the benefits of the technical change. This was done by taking into account the characteristics of demand and supply for the individual input categories. They also showed that trade and other economic policy played an important role in determining how the benefits of the technical change were distributed between consumer and producer groups.

The final contribution to the methodology using economic surpluses was made by Hayami and Peterson (46). Although the model will not be elaborated here, their analysis is important in that it provides a means of evaluating economic research and research of a more general nature than that concerned with production alone. Their study was concerned with estimating the social returns to government expenditures on public information services such as the collection and reporting of information useful for decision making in both the public and private sectors. Their particular interest was the statistical reporting of US farm commodities.

The basic assumption of their model was that erroneous information causes producers to make erroneous production decisions and also distorts optimal inventory carryovers. Hence, marginal improvements in the accuracy of these statistics reduce the social cost of misinformation, which in turn can be considered as an increase in net social welfare. By relating the marginal improvements in the net social welfare to the marginal cost of providing more accurate information, they can estimate marginal social benefit-cost ratios for the various levels of accuracy of the information.

We believe this basic formulation of the problem has more general application. Agricultural research in its broadest sense is in effect nothing more than providing information to relevant decision groups. As the evaluation of research moves beyond its past concentration on output-increasing production research, we believe the Hayami-Peterson model points the way to a more general methodology.

To conclude this discussion of methodologies using the economic surpluses concepts, we note that Lindner and Jarrett (57) have recently made a critique of the various assumptions used by previous authors. Since their paper deals with some of the finer parts of the analyses, we pass over them here.

c. The Production Function Approach

A completely different approach to estimating the payoff to and the effectiveness of agricultural research is to specify the production function for the commodity or the agricultural sector as a whole in a sufficiently broad framework that social inputs such as agricultural research can be included in it. The conventional production function includes only on-the-farm inputs. But inputs provided by the public sector can (and perhaps should) be included just as well.

Griliches (43) was perhaps the first to use this approach, but it has also been used by Peterson (69), Evenson (31) and others (55). The advantages of this approach are that it effectively controls for the other inputs used in production (if the data are measured accurately) and provides an estimate of the marginal product of research. (Methods referred to above all give estimates of the average return.) A marginal return is more useful than an average return to decision-makers studying the merits of new research projects.

An important contribution of the Evenson analysis was to throw light on the time path of response to increased expenditures on research. He found that the returns over time first increased and then decreased, with the high point occurring after about six years. Estimates of the rate of return to expenditures on research are sensitive to the time path in which benefits are forthcoming. The insights gained by the Evenson analysis can be used to improve the estimates of the rates of return.

The production function approach is potentially as rich as the economic surplus approach, although its flexibility lies in somewhat different directions. For example, attempts have been made recently to refine this approach so that it could be used for decision-making purposes. Bredahl and Peterson (16) estimated an aggregate production function for US agriculture using 1969 state data as observations, and included expenditures on agricultural research as an explanatory variable. They were interested in comparing the productivity of research among cash grains, poultry, dairy, and livestock. Their analysis provides estimates of the marginal rate of return to incremental changes in the investment in research on each commodity group, plus an estimate of the marginal rate of return for each state. Hence, their results provide a guide as to productive reallocations of research expenditures both among commodity groups and among geographic areas. 1/ A similar analysis might be done using country observations rather than state observations.

Evenson, Flores, and Hayami (33) have used the basic production function approach to deal with the technology transfer problem. Their interest was in analyzing the costs and benefits of rice research, a case where both an important International Center and national research programs have produced new knowledge which presumably differs in its transferability. They also attempt to take account of knowledge transfer from other disciplines.

The basic model amounts to regressing the change in yield of rice over a base period on a set of farm input variables that would be expected to influence yields plus a set of knowledge stock variables that are constructed as cumulated research investment. The knowledge stock variables include (1) research undertaken in agronomy and plant breeding specifically to improve rice technology; (2) research activity in plant physiology, phyto-pathology, and soil science (research which is not commodity specific); (3) agronomic and plant breeding research activity in countries other than the country in question, but which are in the same geo-climate region; and (4) agriculturally related scientific research in other countries located in the same geo-climate zone.

The regression analysis provides a means of statistically isolating the effects of these various research programs, while at the same time controlling for the use of other inputs that are expected to influence observed yields. The benefits of research can then be imputed to particular research programs, and one has the basis for answering allocatory questions. In principle, increments in budget would be allocated to where the social rate of return were highest.

d. Impact of Research on National Income

Tweeten and Hines (100) have used a still different approach to the evaluation of the returns to agricultural research. Their approach is somewhat

---

1/ Robert Thompson used similar methodology in his Ph.D. dissertation to estimate the marginal rates of returns to research activities by state in Brazil. See Thompson and Schuh (99).

similar to the input-saving methodology, and recognizes explicitly that a contribution of new agricultural technology is the resources it releases to the nonfarm sector.

Tweeten and Himes calculate how much lower the national income would be if the percentage of people on the farm were still the same as in 1910 and the resulting additional farmers had the income of today's farmers instead of today's non-farmers. This provides a measure of the benefits from research. Then they estimate the costs of public and private research, education, and federal programs, and use these to calculate a benefit/cost ratio.

This approach provides only a rather crude approximation to the benefits of agricultural research. It does provide it in a form understandable by policy workers, however, and may be feasible when data are rather scarce.

e. Nutritional Impact

Research goals for agriculture may be expressed in nutritional objectives. Pinstrup-Andersen, Londono, and Hoover (72) have developed a procedure to estimate the nutritional implications of alternative commodity priorities in agricultural research and policy. Their model estimates the distribution of supply increases among consumer groups, the related adjustments in total food consumption, and implications for caloric and protein nutrition. This procedure permits a translation of increases in agricultural output to its impact on nutrition, and by income groups. Hence, equity and nutritional considerations can be analyzed.

This model is rather demanding in terms of detailed knowledge of demand parameters and present consumption patterns. It does not in itself provide estimates of rates of return, but could be extended to such an analysis if nutritional objectives were translated into a suitable form. This approach does provide information that could be of considerable value in establishing research priorities if improved nutrition is the research goal.

* * *

To conclude this section, we would note that a rather rich set of research procedures have been developed whereby research can be evaluated and its contributions and various effects analyzed. It should also be rather obvious that there is no one approach that offers a panacea. Different approaches are useful for answering different questions, and the particular question will vary a great deal depending on the individual problem situation.

The main conclusion we draw is that there is an ample methodological base for using data generated by the economy and past experience to understand and analyze the consequences of investments in agricultural research and its interactions with economic policy and institutional arrangements. The knowledge gained from these analyses can in turn be used to improve the decision-making with respect to scarce research resources.

B. Ex Ante Models

In this section we review some of the major procedures and methods that have been used to improve ex ante decision-making with respect to research resource allocation. The literature in this area is vast, since such procedures are widely used with industrial and military research. 1/ Our review concentrates on the models developed for agricultural research. These models range from approaches which provide a systematic means of utilizing informed judgment, to approaches which attempt to provide empirical knowledge on the consequences of alternative causes of action.

Before discussing some of the methods suggested in the literature, however, it is fitting to recall the words of Cetron and Johnson (22): "We are well aware of many of the omissions and weaknesses of these quantitative-selection or resource-allocation techniques. It should be stressed again that they are not intended to yield decision, but rather information that would facilitate decision. Indeed, these techniques are merely thinking structures to force methodical, meticulous consideration of all factors involved in resource allocation. Data plus analysis yield information. Information plus judgment yield decisions." 2/ And they go on to say: "It is wrong to say that one must select intuitive experience over analysis or minds over machines; really they are not alternatives; they complement each other. Used together, they yield results far better than if used individually."

Pure analysis or pure intuition should not be the objective of any decision-making group. There is an unknown optimal mix of analysis and intuition. This optimum combination cannot be defined a priori and generally. It has to be reached through the joint effort of analysts and managers. A continuing, sequential, dialetical process will be necessary to determine what is necessary and possible with respect to Loasby's (58) "width of agenda", "set of control variables", and "degree of programming." 3/

Decision problems such as those associated with the allocation of resources to agricultural research have been classified according to three major criteria. The first is according to time, with the problems classified according to whether they are static or dynamic. For decision

---

1/ For an analysis of the state of the arts as of 1968, see Rubenstein (78). He calls attention to the fact that up to that time much of the literature was the result of laboratory experiments, not models that were actually implemented, and that to be useful to the research manager the entire process (problem recognition, research, development and testing, engineering and tooling, adoption and continuous improvement) had to be carried out.

2/ Underlined words were stressed in the original.

3/ See below for more detail.

models formulated in a static framework, time is not considered explicitly. The actions and reactions involved in these situations are either assumed to take place instantaneously, or without a time dimension. For models formulated in a dynamic framework, time is considered explicitly and the actions and reactions are assumed to occur at different distinguishable points in time.

There is also a compromise between static and dynamic models, the comparative static framework. In this case a static analysis for different points in time is made, and then a comparison is made of the results over time. The objective is to keep the simplicity of static models while gaining some advantages of the dynamic perspective.

A second criterion according to which decision problems may be classified is according to the degree of uncertainty 1/ involved. At one extreme, models have been built assuming perfect knowledge with respect to all events taking place. It is assumed that the decision-maker has complete certainty about actions and resultant outcomes, both qualitatively and quantitatively. These models are called deterministic. Against these simpler models, there are models that try to take into account the realities of uncertainty facing decision-makers in any area of endeavor. These are called probabilistic models, since they try explicitly and formally to consider the fact that certain variables have a probability distribution of possible values and are not under the direct control of the decision-maker. The most the decision-maker can do is to insure himself against the risk involved in this situation. As with any insurance scheme, premiums and the probability distribution of losses will have to be evaluated. An extreme case of uncertainty is when the probability distribution of the uncertain events is not known. 2/

The third criterion according to which decision problems may be classified is according to the "environment" in which the decision-maker, be it a person or a group of persons, has to take decisions. To the extent that there is not an intelligent and informed force reacting to the decision-maker in a competitive way, i.e., to the extent there is no conflict of interests between the decision-maker's decisions and the opponent force's objectives or goals, the decision problem is represented by a one-decision-maker model. Otherwise, conflict of interest between competing decision-makers, persons, or groups of persons, results in the necessity of the decision model taking into consideration each opponent decision maker's reaction to one's decisions. These much more complicated decision situations are analyzed through models generally designated by two- or n-decision-maker models. 3/

1/ No effort is made here to distinguish carefully between risk and uncertainty in Knight's (54) sense. A discussion of the distinction between risk and uncertainty may be found in Roumasset (76).

2/ A discussion of risk and uncertainty has been part of the literature on decision theory for many years. A standard reference for these concepts and for the discussion of subjective versus objective probability may be found in Luce and Raiffa (59). See also Hampton et al. (44).

3/ Again, Luce and Raiffa (59) is a good reference.

Most of the decision problems in the area of agricultural research resource allocation may be treated as a one-decision-maker situation and therefore do not call for models that represent conflict of interests. But both time and uncertainty are important elements in agricultural research. Moreover, to obtain the possibility of explicitly considering time and uncertainty in analytical models of research resource allocation, a price has to be paid in terms of mathematical complications.

The degree of complexity will also depend on other factors. Loasby (58) classifies decisions according to three aspects: (1) width of agenda, (2) specification of the set of control variables, and (3) degree of programming. The width of agenda refers to the definition of the system boundaries. The decision to consider single or multiple goals, as well as single or multiple restrictions, is a relevant issue in this context. The specification of the set of control variables had to do with the planning horizon or, as it may be put, with the length of the run, whether short or long. The degree of programming depends on how precisely the decision procedures are prescribed. It is clear that in modeling decisions, the degree of difficulty will depend on how broad the system's boundaries are, how many variables are considered simultaneously, and how well programmed the whole procedure is.

a. Scoring Models

Scoring models are relatively simple procedures to formalize the decision process involved in the choice of a research portfolio. Key evaluators, usually the scientists themselves, are called upon to express their evaluation of alternative research projects. These evaluations are based on the potential contribution of each research project to a prespecified goal or set of goals. These goals can be measured in a continuous or discrete way, but evaluations are expressed numerically. Where more than one goal is involved, the same or other evaluators will have to establish a weighting structure. That is, they will have to express numerically the relative importance to society of each goal, especially if they are competitive. Complementary goals can be reduced to only one goal.

Scoring models are generally very flexible not only with respect to the number of goals, but also with respect to the type of goals considered. The specific approaches we will review here are those developed by Iowa State, North Carolina and NASULGC-USDA.

(i) The Iowa Model 1/

The primary purpose of the Iowa model was to ensure the greatest return for the research money spent at the experiment station. 2/ In addition, however, an increase in the value of research output was expected due to better evaluation and selection of projects, and an increase in resources for research due to an improved ability to demonstrate their efficient use.

1/ An algebraic representation of this model is presented in Appendix A.

2/ See Mahlstede (60).

A first and very important step in the implementation of the Iowa procedure was a decision by all administrators involved regarding the necessity for a more formal method of resource allocation, and their agreement on and commitment to a proposed scheme for evaluating research projects. The next step consisted of developing a set of goals and sub-goals. Three goals were selected: growth, equity, and security, which were assumed to apply to the state of Iowa as well as the country as a whole.

The first experiments with the model considered only growth. This was due to the difficulties involved in weighting the three goals. Nevertheless, this review will be concerned with the conceptual model in which all three goals were considered.

The research effort of the station was divided into three major areas: commodity research, resource research, and agricultural management research. These three areas were further subdivided into 19 sub-areas comprising one or two products, or resources, or aspects of agricultural and management research. To each one of these 19 areas a panel of experts was assigned.

The panels were then asked to identify all research alternatives in each area. Each panel member had to present a list of research projects that in his opinion would represent a significant contribution to knowledge and to the goals of growth, defined as value of resources saved and as value of increased output. The panel had to consolidate these individual opinions into a list of suggested research activities. In addition, for each alternative research project suggested by the panel, an estimate of the cost in terms of science man-years and other supporting costs was presented.

As a second step in the procedure, these lists of suggested research activities were submitted to another special panel for evaluation in terms of the second criterion, equity. The panels were asked to give an evaluation of each project in terms of its contribution to absolute and relative equity. This procedure was then repeated in order to take security into account.

A scoring procedure, where each project receives a "grade" on a given scale according to its contribution to growth, and then to equity and security, is the core of the method. The difficulty in scoring is evident, as well as its strong dependency on the experience and wisdom of the panel members.

This is a static, deterministic model insofar as time and uncertainty are not explicitly accounted for. The effect of some kinds of uncertainty can be evaluated by means of sensitivity analysis, however.

A question might be raised as to whether this model requires too much guessing. The answer is probably yes, but they can be informed guesses, which tend to improve with repeated trials. A second question is whether

this model is better than no model at all. Again the answer is probably yes, so long as scientists and decision makers work together to improve it.1/

(ii) The North Carolina Model

The major question addressed in the North Carolina Agricultural Experiment Station model was how much emphasis (in terms of resources) should be put into each of the research problem areas. 2/ The procedure involved several groups of interdisciplinary research and extension faculty, plus several groups of external scientists. Groups of administrators from the experiment station and from the departments were also involved.

The interdisciplinary teams of researchers and extensionists were allocated to the following research problem areas: biological sciences and technology, animals and plants, environment and natural resources, and food-fiber-people-economics. They reviewed the entire research program and prepared recommendations on how and when human and monetary resources should be reallocated. Then they rated each of their recommendations according to criteria such as the extent to which the research met state experiment station, department and national goals, the urgency of the problem, cost, relevance, likelihood that research results would not be available elsewhere, and potential contribution to knowledge. 3/ Not all of the above criteria were used in each of the four areas into which the research program was classified. However, a weighting system was developed for the criteria through use of a Delphi procedure 4/ involving the department heads and the administration of the agricultural research station.

The recommendations of the interdisciplinary in-house teams were then submitted to several smaller groups of external scientists. After reviewing these recommendations, the external groups of scientists developed their own set of recommendations for resource reallocations within the research program, and rated the recommendations of the in-house teams.

The recommendations of the interdisciplinary in-house teams were finally rated according to the criteria discussed above by each member of

---

1/ A version of this model was used in the recent World Food and Nutrition Study conducted by the National Academy of Sciences (64). The individual panels were urged to make quantitative estimates of the expected costs and benefits from the priority lines of research.

2/ See Shumway (89) for a detailed report on the North Carolina procedure. See USDA (102) for a description of the classification of research problem areas.

3/ These criteria were adopted from the criteria recommended in USDA (102). See also Williamson (106).

4/ On the Delphi method see Dalkey and Helmer (25), Beattie and Reader (14), and Brown and Helmer (18).

three groups: in-house scientists, external scientists, and department heads. Research area scores were then computed by an algebraic formula, 1/ averaging over the criteria all the partial scores attributed to the recommended increase in resources to a given area. Weights representing the relative importance of each criterion were used to arrive at a weighted score. These scores represented the evaluation of individual scientists and administrators of the importance of each research area.

The average score was then computed for each research area. This average score was a numerical expression of the "average" opinion of in-house and external scientists, plus administrators, with respect to the "relative worth" of each research area, given the criteria set that was developed independently.

2. The North Carolina method, like other scoring methods, has the advantage of forcing all people involved to spell out formally what they think each research effort will contribute to given goals while at the same time respecting some restrictions. A feature of the North Carolina model that deserves special attention is the computation of two related measures to rank the research areas: the "average" and the "average minus one standard deviation." The "average minus one standard deviation" measure for research areas with the same average score will give preference to the research areas in which opinions were closer together. That is, it will give a higher rank to research areas where consensus is greater.

Another point to note 2/ is the diversity of opinions demonstrated by the North Carolina effort. The degree of (linear) association among scores given to different research areas by any of the three groups of scorers involved was low (the highest was 0.45). The same low association was found within groups of scorers.

The major fault with the North Carolina model, aside from those due to the intrinsic characteristics of scoring models, was the failure to specify the goals more precisely. Consequently, each scorer could have a different idea about the goals of the experiment station, the departments, and the country.

(iii) The NASULGC-USDA Model

The National Association of State Universities and Land Grant Colleges and the USDA jointly defined and implemented a systematic procedure for evaluating and allocating resources to agricultural research. 3/ A task force was assigned to the study and its first step was to prepare a

---

1/ See Shumway and McCracken (91).

2/ This point is discussed by Shumway and McCracken (91).

3/ The basic document reporting this effort is USDA (102). See also Bayley (13). The present discussion closely follows Williamson's (106) excellent presentation.

general scheme for the classification of research. Research was classified into three major headings, according to whether it was related to an activity, a commodity or resource, or a field of science. The short term objective of the classification scheme was to assemble information to describe research programs and to project needs and priorities for future research. The longer term objective was to develop an information storage and retrieval system.

A second task in the general procedure was to classify according to this scheme all research going on at the Department of Agriculture's research units and at the state agricultural experiment stations. Information was also developed on the number of scientist-man-years and funds expended during fiscal year 1965. This permitted an estimation of scientist-man-years and funds being devoted to each research category in the classification schemes.

The third task was to establish national goals for agricultural research. Usually, national goals are too broad for the relationship between them and the agricultural research output to be evident. 1/ It was at this stage of the study that the Planning Programming Budgeting (PPB) 2/ system was adopted by the Department of Agriculture. As Schich (81) says, "One of the major aims of PPB is to convert the annual routine of preparing a budget into a conscious appraisal and formulation of future goal and policies." Then, the study and PPB had to be adapted to each other in terms of the study's goals and the PPB's "missions". 3/

The research program was then divided into six areas: (1) soil, water and air; (2) forestry; (3) horticultural crops; (4) field crops, (5) livestock and poultry; (6) agribusiness and human resources. To each of these six areas a review panel was assigned consisting of people from universities, federal and state agencies, private research organizations, producer groups, industry, and members from the original group responsible for the study. The objective of these review panels was to go into the details of

---

1/ Mahlstede's (60) discussion of the Iowa model presents examples of how to relate social goals like People's Welfare to intermediately defined research lines. His example is not a general solution, though. See also Pinstrup-Andersen and Franklin (71).

2/ According to Puterbaugh (73) the PPB system implied the use of a problem-oriented budget, a "zero-base" budgeting process in place of the usual incremental budgeting, and multi-year budgeting. All these aspects involved difficulties.

3/ It is interesting to note Williamson's (106) comment that "Final agreement on the goals was not reached until the estimate of future research needs was essentially complete."

each area, subdividing them into research problem areas and providing estimates of the number of science-man-years needed in the future. 1/

b. The Minnesota Model

MARRAIS is classified as a multi-dimensional ranking method. 2/ It is a well thought out logical structure that takes into consideration many of the uncertainties involved in the prediction of costs and benefits in research.

Three major steps are involved in MARRAIS: specification, estimation, and analysis. A fourth step would be selection of the research portfolio, but this is not within the scope of MARRAIS itself. Selection is left to the decision-maker. MARRAIS is an information gathering and processing device to help decision-making, just as most of the other existing methods for resource allocation analysis attempt to be.

In the specification phase, the alternative research projects to be analyzed and evaluated are defined under lines of administrative responsibility. (MARRAIS cannot help in identifying the research alternatives.) In addition, the form of the research results and the unit of measurement of these results are specified at this stage. MARRAIS uses percentages of total objective achievement as the unit of measurement. This is to take account of the fact that benefits may be derived from research efforts that do not fulfill their objectives. Another point to be specified in this stage is who is going to provide all the necessary estimates related to research costs and benefits. The choice of "estimates" has no sound theory on which it can be based.

The second phase in MARRAIS is estimation. To understand the estimation phase, it is necessary to understand the logical structure of the model. 3/ MARRAIS works either with a present-value formulation of benefits and costs, deriving from these benefit-cost differences (B-C) and ratios (B/C), or with an internal rate of return (IRR) formula. All the estimates are made under alternative levels of average annual expenditure and expected true horizons for project completion.

MARRAIS recognizes that in practice a lot of uncertainties affect the discounted present value of costs and benefits, so they should be thought of as stochastic variables with given probability distributions. Moreover, the discounted present value of benefits depends on certain variables. MARRAIS hypothesizes that it depends on the annual benefit accruing from the research, assuming 100 percent adoption of its results, on the adoption patterns over time, on the "scrap" value of certain research facilities, and on the so-called "process" value of research (the increase in the value of

---

1/ The "future" at that time was the current 1977 year.

2/ See Shumway (88). For a "best informed" exposition on MARRAIS, on which the present discussion is based, see Fishel (36).

3/ An algebraic representation of the model is presented in Appendix B.

participating scientists plus increased human capital from graduate training involved with the research effort). On the other hand, MARRAIS hypothesizes that the stochastic (discounted) present value of costs depends on the average annual expenditure on research, on the maximum annual expenditure on dissemination, and on the time path of dissemination costs of research starting one period after the project is completed. It should be noted that both the adoption patterns over time and the time path of dissemination costs are functions of time.

Groups of experts 1/ provide estimates for the variables. Given an average annual expenditure on a project, they estimate the probability of the project being completed in alternative periods of time. Then, with estimates of the mean time to complete the project and the average annual expenditure on the project, benefits are estimated. This results in a probability distribution of benefits from a given project, funded at a given level of annual expenditure and taking the mean expected time for completion. This probability distribution of benefits is weighted by the probability distribution of technological feasibility for each project and funding level. By a random sampling from the distributions of the involved stochastic variables, plus single-value estimates of the non-stochastic variables, estimates are made of the distributions of the difference between benefits and costs, of the benefit-cost ratios, and of the internal rate of return.

MARRAIS is a relatively sophisticated, multiple-dimension ranking model. There may be serious difficulties in applying it in an international context due to the large degree of variation to be found in the relevant variables and the difficult task "estimators" would face. On the other hand, these same conditions are an argument in favor of a model such as MARRAIS. If the difficulties are recognized, an attempt can be made to deal with them in a systematic way. Moreover, the potential for sensitivity analysis which MARRAIS offers with respect to all these uncertain phenomena may be very useful. The use of different parameters for the distribution of the stochastic variables, or the use of different probability functions, are some of the ways of dealing with the precarious confidence that is usually put on such "subjective" distributions. But it may be very expensive both in time and money.

c. Pinstrup-Andersen and Franklin Model

The model developed by Pinstrup-Andersen and Franklin (71) is an attempt to reflect their argument that "concurrence between the technology specification received by the scientist and the technology which results in maximum contribution to the achievement of social goals is the responsibility of research management". This concern leads them to analyze the problem of defining working objectives for research from stated national development goals. It should be kept in mind that they are concerned with the allocation of agricultural research resources in developing countries. This helps explain the greater emphasis put on the relationship between development goals and agricultural research relative to other studies.

1/ See Shumway (88) for a discussion of group procedures.

The authors argue that after the identification of the changes in product supply, input demand, and farm consumption necessary to attain the development goals, the identification of the research problems should be made independently of the alternative technologies that can contribute to the solution of the problem. They call this a "technology-free" specification of the problem, since it does not presume ex ante a particular technological solution to the problem.

This is a very important aspect of the problem definition phase in scientific research. It relates to the "form" in which the problem is presented. The advice is not to jump from farm problem identification to research problem definition in terms of required technology without a careful evaluation of all technological solutions available. Another point to keep in mind is that definition of the problem at the farm level is not an easy task. The work of Hayami and Ruttan (47) makes it clear that low production per acre in one region relative to another region does not necessarily imply a problem in the first region. The differences in productivity may reflect a difference in factor proportions, which in turn are induced by differences in relative factor prices.

After the identification of problems in a "technology-free" manner and, subsequently, of the alternative technologies available to solve each and every problem, it is necessary to estimate the time, costs, and probabilities involved in research and in farm adoption for each alternative technology. The next step is to estimate the impact of the research alternatives on farm consumption, product supply, and input demand. This effort requires some previous knowledge about the economic structure (parameters) of the relevant production sectors and of the product and input markets.

Then the estimated effects on farm consumption, on product supply, and input demand are used to obtain an evaluation of the contribution of the alternative research lines to the achievement of the development goals. This leads to a specification of the working objectives for the research and of the desired technology.

The representation of this approach by means of mathematical equations is not simple, since it is a system approach to the allocation of research resources. Moreover, it tries to relate specific research problems to overall aspects of growth, equity, and security through consideration of variables such as income distribution, nutrition, demand for labor and other services, farm consumption, capital formation, supplies and demands, net revenues, risk, etc. A flow diagram usually gives an easier pictorial understanding of the interrelationships involved, and such a flow diagram can be found in Pinstrup-Andersen and Franklin (71, p. 424).

Eight social goals are considered, with the objective being to obtain an improvement in the following variables: growth, income distribution, employment, farmers' net income, farmers' cash inflow, human nutrition, degree of self-sufficiency in food, and foreign exchange earnings. Other goals might be considered, additionally and/or in substitution of the eight cited, but the interrelationships and parameters in the model will have to be modified accordingly.

This model can be applied to the problem of resource allocation in single product or production factor research (as most of the centers in the international agricultural research system are organized), multi-commodity research, and farming systems research (such as with small farms). Some promising empirical results have been reported by Pinstrup-Andersen and Franklin.

Pinstrup-Andersen and Franklin's model is an example of a useful effort in the area of research on research, and deserves further testing and development. It also involves a relatively complicated methodology, however, if research management does not want to invest significant human resources in the problem of resource allocation. As a guide to the kind of information useful to the analysis of resource allocation, the model is quite useful. Its claims that the effects of research are highly dependent on the kind of public policy being pursued by the country is important and in line with points made earlier in this paper.

d. Cartwright Model

The model developed by Cartwright (20) focuses on the allocation decisions for research within a department of agricultural economics. Two decision problems are analyzed: the choice of research areas to work in and the choice of a research job portfolio.

A large quantity of information is fed into the model of research and decision. The information includes (1) the amount of time each researcher has to input into each alternative procedure available to undertake the new research areas in a given period of time (quarter); (2) the amount of time each researcher could divert from previous assignments to a new category of employment in a given quarter; (3) the amount of funds of a given category that the new research area would bring to the department in a given quarter if a given procedure were adopted; and (4) the number of new staff positions in a given category that would be created in a given quarter if a certain procedure were adopted. With this information and a formal statement of the staff preference function, the allocation problem is cast in the form of a non-linear integer programming problem.

Generally, information in (1) and (2) above is not readily available and thus requires some additional estimation by the staff. Information in (3) and (4) is generally available in any proposal for undertaking a new research area or for reallocating resources among current research areas.

The job portfolio selection model assumes that a more centralized decision procedure is implemented in the academic environment. It involves a simultaneous evaluation and selection of all research jobs once each quarter. Research is classified into four classes according to whether they are dissertion research or not, and whether they are in the agricultural experiment station program or are funded by outside research contracts. A fifth class, of currently active jobs, is also explicitly included, subject to a decision of continuing or stopping these jobs. A job can be modified with respect to time horizons, resource use, etc., by terminating it and at the same time starting a job with the required modifications.

The information required for the model includes (1) the amount of time required from each staff member by each alternative job in a given quarter; (2) the annual expenditure, by class of expenditure, from each budget component, required by each alternative job; (3) the amount of time uncommitted, by researcher and quarter; (4) and the annual budget, by expenditure class and by budget component. Again, staff preferences have to be evaluated in a tentative effort to define the goals and, eventually, an objective function. This is an important and difficult task in the procedure.

The models for research-area decisions and for job-portfolio decisions are structurally very similar. Cartwright (20, p. 151) calls attention to the following five common characteristics: (1) the decision variables are interdependent and linearly related, (2) these variables cannot assume negative values, (3) at least one variable can take on only one of two values and hence is not continuous, (4) the objective function contains several goals (some of which may not have a natural or obvious measurement) with an unknown mathematical form, and (5) the model can assume certainty about all facts involved or it can assume that risks and uncertainties are present.

The solution procedure for the model is not straightforward. An optimization procedure may be used, as well as one that although not finding an optimal solution, uses stimulation to indicate some alternative, acceptable solution.

Cartwright's model is very imaginative. However, this is an example where more "development" of the model would be necessary before its use in routine decision making would be practical. Further development of the goals-preferences-objective function procedure is necessary, as well as of the solution procedures. Also, since it was cast in the framework of a department of a univeristy, a large number of modifications would be necessary to adjust this model to the International Agricultural Research Centers' conditions and needs.

A short test of a research monitoring and reporting system (REMORS) is also reported by Cartwright (12). This system consists of a register of all current and planned research, plus a system of reporting research progress. Reviews of the "cost-benefit" of REMORS by staff members were mixed. Of course, they probably had a very short-run review of the procedure. A long-run view, taking account of expected improvements due to continued testing and development, might be more favorable.

The general conclusion we reach is that although Cartwright's work is to be commended for its rigorous and comprehensive focus on quantative methods for resource allocation, it did not go very far into the development stage, even in the special environment of an academic department for which the model was designed. Therefore, it does not seem appropriate for the international agricultural research system at this moment.

e. <u>Castro and Schuh Model</u>

The major characteristic of the Castro and Schuh model is the emphasis it gives to both growth and distributional effects of research

and the resultant technological change. 1/ Using the concepts of producer's and consumer's surplus, functional distribution of income, and two-sector models of general equilibrium, Castro and Schuh use analysis and information on key parameters to assess the effects of research and technological change on given products of a given country. 2/ They establish four goals for a research program: (1) to increase the net income of the agricultural sector, (2) to increase employment and income of workers in the agricultural sector, (3) to increase consumer welfare through lower real food prices, and (4) to maximize the contribution of the agricultural sector to the growth of the overall economy.

Castro and Schuh's model is not a formal, mathematical model, although it could be used as a starting point to build such a mathematical model. 3/ The important point to note is its focus on both the growth and distributional effects of technological change, and on both the direct and indirect effects of research. 4/

The explicit consideration of the distribution of benefits and losses between producers and consumers and among factors of production is a desirable characteristic of Castro and Schuh's procedure. 5/ Also important is the following up of the effects and reactions of technological change in the agricultural sector into the non-agricultural sector.

This model minimizes the burden put on scientists and administrators in terms of the amount of information necessary and in terms of the amount of difficult estimates and/or "informed guesses" required. The model depends primarily on secondary data, and the burden of the analysis rests with the analysts.

1/ This model is discussed in Castro and Schuh (75). Details may be found in Castro (74).

2/ Cotton, sugar cane, rice, corn, edible beans, and manioc in Brazil.

3/ It is worth noting that Pinstrup-Andersen and Franklin's (71) model is in a sense a formalized version of Castro and Schuh's analysis. Of course, it is formally much larger and more comprehensive, with subsystems for the demand side cast in terms of matrices of elasticities, etc.

4/ Brinegar (17) argues strongly in favor of taking account of indirect (second and higher order) effects of technological change through long term multipliers. Quoting another author in the area of education to the effect that "recent estimates of high economic return to education could be very misleading because the real returns are much higher," Brinegar implies that the same might be occurring with research because secondary effects are not followed through time.

5/ Bayley (13) in addition to warning against a "Greek philosopher's attitude" with respect to new methods for research resource allocation, calls attention to the need to better identify how benefits and losses are distributed.

A point to note is that in Castro and Schuh's model no explicit treatment of the probabilities of success (technical, "commercial" or other) is presented. This is as difficult an aspect of ex ante models as it is important. The question involves how much difference it makes to discriminate even grossly among probabilities of success for alternative research endeavors relative to not discriminating at all. Not discriminating implies that for all practical purposes the probabilities are assumed to be equal. The question that arises is whether it is more desirable to give an equal probability of success to all alternative research projects than to try to discriminate among them. Ultimately, this is an empirical question. The answer may have to wait for further research.

Another point worth noting is that the Castro and Schuh model was developed within the framework of a given country, with given goals, institutions. policies and endowments. The international agricultural research system, even in the case of a Center concerned with only one commodity, operates in a much wider milieu. The task of following the distributional effects among all classes of exporters, importers, producers, consumers, regions, sectors, etc. becomes much greater. However, this problem is present whatever the allocation model chosen. The fact that some models dodge this obstacle by not giving any consideration to these aspects does not make them any more desirable.

f. Easter and Norton Model

Easter and Norton (29) apply an ex ante benefit-cost analysis to requests for additional resources in the federal budget of the Land Grant Universities. Specifically, they consider the case of certain research program areas in corn and soybeans in the North Central region. They also discuss applications of benefit-cost analysis to livestock and to rural development research.

In discussing the criticisms that have been made against the use of benefit-cost ratios in ex ante analysis of research resource allocation, they argue that "while problems of estimating benefits preclude the determination of an 'optimal' allocation of research resources, quantitative cost benefit techniques may help policy makers improve their decisions. Certainly as a minimum, carefully calculated estimates of benefits can be compared with costs to determine which projects will likely yield positive returns."

Benefit-cost ratios are calculated using the low side of a range of estimates provided by scientists on the yield and cost effects of each research line, as well as on the expected adoption rates for the new technology. Also, a discount rate is adopted (10% in the example) together with product prices and probabilities of success for each research alternative. Some assumption is made about the trend in area cultivated with the crop, in the present example that it would remain constant at the base year (1975) level. Product quality was also assumed either unchanged or, if improving, not affecting the cost of livestock feeds. Benefits were followed through a period of 25 years ending in the year 2000.

An important aspect of the analysis of required additional funds for research in the Easter and Norton approach is the sensitivity of benefit cost ratios to single and combined variations in the probabilities of success, the expected yield increases, the product prices, and the length of the lag between research expenditures and the availability of the research results to the farmers. While it is easy to anticipate the direction of the changes in the benefit cost ratios due to changes in these conditions, the careful application of sensitivity analysis to any real, quantified evaluation, is a very important source of information to the decision-makers. As a matter of fact, it works as a kind of feedback process to let administrators know the relative importance of added precision and accuracy in the estimation of the several variables involved in the evaluation.

An effort was also made to assess the distribution effects of the research program. The increased-production effect of the research was spread over the related sectors of the economy. In the example used by Easter and Norton, the effect of a 3 percent increase in the production of corn and soybean on the feed, livestock, and meat economy was traced through. The price elasticity of demand for corn and for soybeans were used to assess the effect of increased production on farm gross income. These elasticities appear to be close to unity, since gross farm income is estimated to remain unchanged. Effects on the soybean oil market were also analyzed, as well as the long run effect on the livestock sector.

In evaluating research projects in livestock, the benefit-cost ratio, given the necessary adaptations, was still considered useful, in spite of a greater difficulty in assessing benefits. In the rural development research and extension project appraisal, Easter and Norton turn to cost effectiveness as an operational method. They call for information on (1) the research and extension objectives, (2) a cost estimate by objective, and (3) expected outcomes in physical (in money if possible) terms. Alternative ways of reaching the same results are then assessed.

As one of their conclusions, Easter and Norton stress the key role that the cooperation between scientists and social scientists plays in the effort to evaluate ex ante research efforts.

8. Atkinson-Bobis Model

Simulation, viewed in a broad perspective to include several random sampling procedures, has been utilized in several models for research resource allocation. In the studies reviewed above, simulation was normally used in the approaches of Fishel, Pinstrup-Andersen and Franklin, and Cartwright. 1/ Simulation is at the center of Atkinson-Bobis' model.

1/ In 1972, Souder (93) estimated that much more than one hundred models of research resource allocation had been built. Simulation procedures must have been more or less an important component of the solution approach to a good number of these models.

The model developed by Atkinson and Bobis (8) for the industrial sector has fared well in the opinion of experienced research managers and of model builders. 1/ It is not a pure simulation model since it involves optimization through dynamic programming. A random sampling procedure is used to take account of the stochastic nature of some of the variables involved in the procedure, producing a distribution of losses and of positive returns, in dollar terms, around the expected returns.

The Atkinson-Bobis model has a probability component relating expenditures and probability of successful completion of each project, in each year and after any number of years within the planning horizon (five years of investments and eleven years in total for the product life in their example). Probabilities of technical, legal, engineering, and commercial success are established, and the product of these probabilities, provided the project is completed, gives the overall probability of success.

Sales estimates for specified years are made and they are fitted into a logistic equation due to the observed fact that new product sales usually follow an S-shaped pattern. The sales function recognizes that the initiation of sales later than anticipated usually results in a smaller share of the market being gained after sales are stabilized. The penalty for starting sales later than in the first year after successful completion of the project is made severe in the Atkinson-Bobis model.

Some assumptions are made about selling prices in each year, as well as about manufacturing costs (profit margins may be used). Overhead and selling costs are assumed a constant proportion of sales revenue.

The revenue from sales in any given year, considering sales started in any previous (or current) year, is then computed. The present value of the revenue stream net of overhead, selling, and manufacturing costs provides an estimate of the net revenue for any given project assuming sales started in a given year. Weighting these values for differences in the starting year of sales with the probability of success of the project in each year gives an expected payoff.

An estimate of the expected discounted research expenditure is obtained by considering expenditures and probabilities in both cases of project failure and success in each year. Then, the expected payoff minus the expected expenditures provides a measure of the expected profit of the project.

The model also has a mechanism to analyze the rate of expenditure over time. Using concepts of efficient research expenditures, effective research expenditure, and the relationship between expenditures in one year and in the following years, an equation is derived to permit adjustment

---

1/ Souder (93) reports that the Atkinson and Bobis model came out well in a comparison of 26 models according to realism (weight = 4), flexibility (weight = 3), capability and facility of use (weight = 2, each), and cost (weight = 1). Shumway (88) calls attention to this comparison.

in efficient expenditures in each year. The mechanism penalizes "over-investment" in the beginning year of the project, due to the higher probability of wasting resources. An optimization procedure is then utilized for the profit variable over a set of projects subject to a budget constraint.

Results from an example presented in Atkinson and Bobis indicated that resources should be concentrated in fewer projects than the usual research policy was funding. An iterative procedure was used to reduce the optimization solution by the number of years, going over it year by year and repeating the allocation process until only negligible improvement in profits was verified.

Atkinson-Bobis resort to a random sampling procedure since the optimization works with point estimates for all the variables involved and since an interval estimation is more secure and informative. Moreover, sensitivity analysis with the optimization procedure would be laborious and not recognize the different probabilities associated with the different values in the relevant range of some of the variables. This simulation through a random sampling methodology permits the generation of a distribution of results (both losses, in the negative range of profits, and positive returns) around the expected profit for each project (the expected profit unbiasedly estimated by the average profit of all the simulation runs for each project).

This is just one example of a model using a kind of simulation procedure. The model was designed to represent the conditions and environment of an industrial concern, and would have to be adjusted for applications in agriculture. However, the form of the final output and some characteristics of the structure of the model make it a necessary consideration when evaluating alternative models.

* * * * * *

The ex ante models described above are but a few examples of the large number of models that have been tried in actual applications. The degree of methodological sophistication ranges from the simpler scoring models to the more complex mathematical programming models.

The advantages of these models are that they provide a basis for decision-making with an eye to the future rather than on what has happened in the past, they pool information from a large number of qualified experts, and they provide a means of relating the research effort explicitly to a set of goals. The disadvantages are that those methods which draw on the opinion of a large number of experts can be quite costly and time-consuming, and that the pooling of a large number of opinions may do little more than pool ignorance. It is probably for these reasons that the more complicated methods have often been used only once.

It is our judgment that one of the methods that seeks to obtain the informed judgment of a large group of experts may have some merit if they are used, say, once every five years. If used in this way, they can provide a source of ideas and suggestions, as well as informed judgment about the appropriateness of research strategies being used. On a continuing basis, we believe there is considerable merit in the approaches taken by Castro and Schuh and by Norton and Easter, since they provide a means of feeding some rigorous analytical research into the decision-making process.

An important need is for more development work with all the ex ante models. In our view the D in R and D for these models has been slighted.

References for Chapter VIII

The full set of references listed in the Schuh and Tollini paper is reproduced for the convenience of the reader. Suggested readings more relevant to the topic of the workshop are listed at the end of this document.

1. Addison, L. E., J. W. Litchfield, and J. V. Hansen, "Managing Growth and Change in an R & D Organization: the Role of Dynamic Modelling," R & D Management, Vol. 6, No. 2, 1976, pp. 77-80.

2. Agriculture and Rural Development Department, "Technology and Agricultural Innovation: An Issues Paper," The World Bank, Washington, March 1977 (mimeographed).

3. Akino, M. and Y. Hayami, "Efficiency and Equity in Public Research: Rice Breeding in Japan's Economic Development," American Journal of Agricultural Economics, Vol. 57, No. 1, February 1975, pp. 1-10.

4. Allen, D. H. and T. F. N. Johnson, "Optimal Selection of a Research Project Portfolio Under Uncertainty," Chemical Engineering (London, No. 241, 1970.

5. Allen, J. A., "A Philosophy on Research Management," Paper presented to National Workshop on Research Management, Melbourne, November 1976.

6. Allen J. M., "A Survey into the R & D Evaluation and Control Procedures Currently Used in Industry," Journal of Industrial Economics, Vol. 18, No. 2, pp. 161-181.

7. Ardila, V. J., "Rentabilidad Social de las Inversiones en Investigacion de Arroz en Colombia," MS dissertation, ICA, Universidad Nacional, Bogota, 1973.

8. Atkinson, A. C. and A. H. Bobis, "A Mathematical Basis for the Selection of Research Projects," IEEE Transactions on Engineering Management, EM-16, No. 1, February 1969, pp. 2-8.

9. Augood, D. R., "A Review of R & D Evaluation Methods," IEEE Transactions on Engineering Management, EM-20, No. 4, November 1973, pp. 114-120.

10. Ayer, H. W. and G. E. Schuh, "Social Rates of Return and Other Aspects of Agricultural Research: The Case of Cotton Research in Sao Paolo, Brazil," American Jorunal of Agricultural Economics, Vol. 54, No. 4 (Part I), November 1972, pp. 557-569.

11. Baker, N. R., "R & D Project Selection Models: An Assessment," R & D Management, Vol. 5 (Special Issue on Project Selection), 1975, pp. 105-111.

12. Barletta, A. N., "Costs and Social Benefits of Agricultural Research in Mexico," Ph.D. Thesis, University of Chicago, Chicago, Illinois, 1971.

13. Bayley, N. D., "Research Resource Allocation in the Department of Agriculture," in Resource Allocation in Agricultural Research, edited by W. L. Fishel, University of Minnesota Press, Minneapolis, 1971.

14. Beatie, C. J. and R. D. Reader, "Quantitative Management in R & D," Chapman and Hall Ltd., London, 1971.

15. Bobis, A. H., T. F. Cooke, and J. H. Paden, "A Funds Allocation Method to Improve the Odds for Research Success," Research Management, Vol. 14, No. 2, (March 1971), pp. 34-49.

16. Bredahl, M. and W. Peterson, "The Productivity and Allocation of Research: U.S. Agricultural Experiment Station," American Journal of Agricultural Research, Vol. 58, No. 4 (Part I), pp. 684-692.

17. Brinegar, G. K., "Discussion: Evaluation and Planning of Research in the Experiment Station," American Journal of Agricultural Economics, Vol. 50, No. 5, December 1968, pp. 1161-1162.

18. Brown, B. and O. Helmer, "Improving the Reliability of Estimates Obtained from a Consensus of Experts," Report P2986, The Rand Corporation, California, 1964.

19. Bush, G. P., "Bibliography on Research Administration, Annotated," Washington, D. C.: University Press, 1964.

20. Cartwright, R. W., "Research Management in a Department of Agricultural Economics," Ph.D. Dissertation, Department of Agricultural Economics, Purdue University, 1971.

21. Caty, G., G. Drilhon, G. Ferne, and S. Wald, "The Research System: Comparative Survey of the Organization and Financing of Fundamental Research," 3 Volumes, OECD, Paris, 1972.

22. Cetron, M. J. and J. N. Johnson, "Technological Forecasting in A Dynamic Environment," in Quantitative Decision Aiding Techniques for Research and Development Management, edited by M. J. Cetron, H. Davidson, A. H. Rubenstein, Gordon and Breach, Science Publishers, Ince., New York, 1972.

23. Cetron, M. J., J. Martino, and L. Roepeke, "The Selection of R & D Program Content--Survey of Quantitative Methods," in Quantitative Decision Aiding Techniques for Research and Development Management, edited by M. J. Cetron, H. Davidson, A. H. Rubenstein, Gordon and Breach, Science Publishers, Inc., New York, 1972.

24. Clarke, T. E., "Decision Making in Technologically Based Organizations: A Literature Survey of Present Practice," IEEE Transactions on Engineering Management, EM-21, No. 1, February 1974, pp. 9-23.

25. Dalkey, N. and O. Helmer, "An Experimental Application of the Delphi Method to the Use of Experts," Management Science, Vol. 9, April 1963, pp. 458-467.

26. Dalrymple, D. G., "Global Agricultural Research Organization," World Food and Nutrition Study: The Potential Contribution of Research, Supporting Papers, National Academy of Sciences, July 1977, Vol. V, Study Team 14, Subgroup B (in press).

27. Dean, B. V., "Evaluating, Selection and Controlling R & D Projects," American Management Association, New York, Research Study 89, 1968.

28. Dean, B. V. and L. A. Roepcke, "Cost Effectiveness in R & D Resource Allocation," IRE Transactions on Engineering Management, Vol. EM-11, December 1972.

29. Easter, K. W. and G. Norton, "Estimates of Potential Returns from Added Research Budget for and Land Grant Universities," Staff Paper, Department of Agricultural and Applied economics, University of Minnesota, November 1976.

30. Economic Research Service, Technology Assessment: Proceedings of an ERS Workshop, April 20-22, 1976, AGERS-31, US Department of Agriculture, Washington, D. C., 1977.

31. Evenson, R. E., "The Contribution of Agricultural Research to Production," Journal of Farm Economics, Vol. 49, December 1967, pp. 1415-1425.

32. Evenson, R. E., "Gains and Losses from Agricultural Technology," University of the Philippines, November 1975 (mimeographed).

33. Evenson, R. E., P. M. Flores and Y. Hayami, "Costs and Returns to Rice Research," Resource Paper No. 11, Conference on Economic of New Rice Technology, IRRI, Los Banos, Philippines, December 1976 (mimeographed).

34. Evenson, R. E. and Yoav Kislev, Agricultural Research and Productivity, New Haven and London, Yale University Press, 1975.

35. Falcon, W. P., "The Green Revolution: Generations of Problems," American Journal of Agricultural Economics, Vol. 52, 1970, pp. 698-710.

36. Fishel, W. L., "The Minnesota Agricultural Research Resource Allocation Information System and Experiment," in Resource Allocation in Agricultural Research, edited by W. L. Fishel, Minnesota University Press, Minneapolis, 1971.

37. Gear, A. E., "A Review of Some Recent Developments in Portfolio Modeling in Applied Research and Development," IEEE Transactions on Engineering Management, EM-21, No. 4, November 1974, pp. 119-125.

38. Gear, A. E., "A Review of Some Recent Developments in Portfolio Modeling in Applied Research and Development, R & D Management, Vol. 5 (Special Issue on Project Selection), 1975, pp. 59-65.

39. Gear, A. E., A. G. Lockett and A. W. Pearson, "An Analysis of Some Portfolio Selection Models for Research and Development," IEEE Transactions in Engineering Management, Vol. EM-18, May 1971.

40. Gotsch, C. H., "The Green Revolution and Future Developments in Pakistan Agriculture," in Rural Development in Bangladesh and Pakistan (ed. by R. D. Stevens, et. al.), University press of Hawaii, Honolulu, 1976, p. 361).

41. Gray, J. A. and S. P. Wolff, "The Changing Face of Research and Development in the Public Sector, R. & D Management, Vol. 6, No. 3, pp. 105-108.

42. Griliches, Zvi, "Research Costs and Social Returns: Hybrid Corn and Related Innovations," Journal of Political Economy, Vol. 66, October 1958, pp. 419-431.

43. Griliches, Zvi, "Research Expenditures, Education, and the Aggregate Agricultural Production Function," American Economic Review, Vol. 54, December 1964, pp. 961-974.

44. Hampton, J. M., P. G. Moore and H. Thomas, "Subjective Probability and its Measurement," in Concepts and Applications of Modern Decision Models, MSU Business Studies, East Lansing, Michigan, 1976.

45. Hayami, Y. and R. W. Herdt, "The Impact of Technological Change in Subsistence Agriculture on Income Distribution," American Journal of Agricultural Economics, Vol. 59, No. 2, May 1977, pp. 245-256.

46. Hayami, Y. and W. Peterson, "Social Returns to Public Information Services: Statistical Reporting of U. S. Farm Commodities," American Economic Review, Vol. 62, No. 1, March 1972, pp. 119-130.

47. Hayami, Y. and V. W. Ruttan, Agricultural Development, An International Perspective, The Johns Hopkins Press, Baltimore, 1971.

48. Hertford, R. and A. Schmitz, "Measuring Economic Returns to Agricultural Research," in Resource Allocation and Productivity in National and International Agricultural Research, edited by T. A. Arndt, D. G. Dalrymple, and V. W. Ruttan, University of Minnesota Press, Minneapolis, 1977.

49. Hertford, R., J. Ardila, A. Rocha and C. Trujillo, "Productivity of Agricultural Research in Colombia," in Resource Allocation and Productivity in National and International Agricultural Research, edited by T. M. Arndt, D. G. Dalrymple and V. W. Ruttan, University of Minnesota Press, Minneapolis, 1971.

50. Johnston, B. F. and J. Cownie, "The Seed-Fertilizer Revolution and Labor Force Absorption," American Economic Review, September 1969.

51. Keeling, A. G., "Comments on Dr. Simmonds' Paper, Costs and Benefits of an Agricultural Research Institute," R & D Management, Vol. 6, No. 1, pp. 43-44.

52. Kislev, Y., "The Economics of Agricultural Research, Some Recent Findings," Food Policy, May 1977, pp. 148-156.

53. Klein, K. K., "An Investigation into the Evaluation of Applied Research Projects in Agriculture," unpublished Ph.D. Dissertation, Purdue University, 1976.

54. Knight, F. H., "Risk, Unvertainty, and Profit," Houghton Mifflin Company, New York, 1921.

55. Latimer, R. G. and D. Paarlberg, "Geographic Distribution of Research Costs and Benefits," Journal of Farm Economics, Vol 47, May 1965, pp. 234-241.

56. Lindner, R. K., "Autonomy andAccountability in Planning Research: The Challenge for Research Management," Economics Department Paper No. 5, the University of Adelaide, South Australia, September 1976.

57. Lindner, R. K. and F. G. Jarret, "Measurement of the Level and Distribution of Research Benefits," Discussion Paper, Economics Department, University of Adelaide, September 1975 (mimeograph).

58. Loasby, B. J., "Choice, Complexity, and Ignorance--An Enquiry Into Economic Theory and teh Practice of decision-Making," Cambridge University Press, 1976.

59. Luce, R. D. and H. Raiffa, "Games and Decisions--Introduction and Critical Survey," John Wiley and Sons, Ince., 1967.

60. Mahlstede, J. P., "Long-Range Planning at the Iowa Agricultural and Home Economics Department Station," in Resource Allocation in Agricultural Research, edited by W. L. Fishel, University of Minnesota Press, 1971.

61. Moore, R. L., "Methods of Determining Priorities in a Programme of Research," R & D Management, Vol. 5 (Special Issue on Project Selection), pp. 59-65.

62. Moore, J. R. and N. R. Baker, "Computational Analysis of Scoring Models for R. & D Project Selection," Management Science, Vol. 16, No. 4, December 1969.

63. Mottley, C. M. and R. D. Newton, "The Selection of Projects for Industrial Research," Operation Research, Vol. 7, November-December 1959, pp. 740-751.

64. National Research Council, World Food and Nutrition Study, Washington, D. C., National Academy of Sciences, 1977.

65. Nguyen, D., "Intersector-Distributional Implications of Agricultural Progress in an Open Economy: An Extension," American Journal of Agricultural Economics, May 1977, p. 374.

66. Nutt, A. B., "An Approach to Research and Development Effectiveness," IEEE Transactions on Engineering Management, Vol. EM-10, September 1965.

67. Paulsen, A. and D. R. Kaldor, "Evaluation and Planning of Research in the Experiment Station," American Journal of Agricultural Economics, Vol. 50, 1968, pp. 1149-1162.

68. Pearson, A. W. "Project Selection in an Organization Context," R & D Management, Vol. 5 (Special Issue on Project Selection), 1975, pp. 59-65.

69. Peterson, W. L., "Return to Poultry Research in the United States," Journal of Farm Economics, Vol. 49, August 1967, pp. 656-669.

70. Pinstrup-Andersen, P. and F. Byrnes (eds.), Methods for Allocating Resources in Applied Agricultural Research in Latin America, Centro Internacional de Agricultura Tropical, Cali, 1975.

71. Pinstrup-Andersen, P. and D. Franklin, "A Systems Approach to Agricultural Research Resource Allocation in Developing Countries," in Resource Allocation and Productivity in National and International Agricultural Research, edited by T. M. Arndt, D. G. Dalrymple and V. W. Ruttan, University of Minnesota Press, Minneapolis, 1977.

72. Pinstrup-Andersen, P., N. R. de Londno and E. Hoover, "The Impact of Increasing Food Supply on Human Nutrition: Implications of Commodity Priorities in Agricultural Research and Policy," American Journal of Agricultural Economics, Vol. 58, No. 2, May 1976, pp. 131-142.

73. Puterbaugh, H. L., "An Appliation of PPB in the Agricultural Research Service," in Resource Allocation in Agricultural Research, edited by W. L. Fishel, University of Minnesota Press, Minneapolis, 1971.

74. Ramalho de Castro, J. P., "An Economic Model for Establishing Priorities for Agricultural Research and a Test for the Brazilian Economy," Ph.D. Dissertation, Department of Agricultural Economics, Purdue University, 1974.

75. Ramalho de Castro, J. P. and G. E. Schuh, "An Empirical Test of an Economic Model for Establishing Research Priorities: A Brazil Case Study," in Resource Allocation and Productivity in National and International Agricultural Research, edited by T. M. Arndt, D. G. Dalrymple and V. W. Ruttan, University of Minnesota Press, Minneapolis, 1977.

76. Roumasset, J. A., "Rice and Risk--Decision Making Among Low-Income Farmers," North-Holland Publishing Company, 1976.

77. Rubenstein, A. H., "Economic Evaluation of Research and Development: A Brief Survey of Theory and Practice," Journal of Industrial Engineering, Vol. 17, 1966, pp. 615-620.

78. Rubenstein, A. H., "Research on Research: The State of the Arts in 1968," Research Management, Vol. 11, No. 5, September 1968, pp. 279-304.

79. Russell, E. J., "R & D Evaluation--The Total Program Approach," Financial Executive, London, No. 234, 1969.

80. Saunders, G. S., "Investment Decisions in Research and Development," Chemical Engineering, London, No. 234, 1969.

81. Schich, A., "The Road to PPB: The Stages of Budget Reform," in Resource Allocation in Agricultural Research, edited by W. L. Fishel, University of Minnesota Press, Minneapolis, 1971.

82. Schmitz, A. and G. Seckler, "Mechanical Agriculture and Social Welfare: The case of the Tomato Harvester," American Journal of Agricultural Economics, Vol. 52, No. 4, November 1970, pp. 569-578.

83. Schuh, G. E., "Some Economic Considerations for Establishing Priorities in Agricultural Research," presented at OLAC Meeting of Program Advisors in Agriculture, Mexico City, November 1972 (mimeograph).

84. Schuh, G. E., "The Exchange Rate and U.S. Agriculture," American Journal of Agricultural Economics, Vol. 56, 1974, pp. 1-13.

85. Schultz, T. W., The Economic Organization of Agriculture, New York, McGraw-Hill, 1953.

86. Schultz, T. W., "What Are We Doing to Research Entrepreneurship?" Agricultural Economics Paper No. 77:11, University of Chicago, March 1977 (Mimeograph).

87. Scobie, G. M. and R. Posada T., The Impact of High-Yielding Rice Varieties in Latin America, With Special Emphasis on Colombia, Series JE-01, CIAT, Cali, 1977.

88. Shumway, C. R., "Allocation of Scarce Resources to Agricultural Research: Review of Methodology," American Journal of Agricultural Economics, Vol. 56, No. 4, November 1975, pp. 716-718.

89. Shumway, C. R., "Models and Methods Used to Allocate Resources in Agricultural Research: A Critical Review," in Resource Allocation and Productivity in National and International Agricultural Research, edited by T. M. Arndt, D. G. Dalrymple and V. W. Ruttan, University of Minnesota Press, Minneapolis, 1977.

90. Shumway, C. R. and J. D. Hwang, "Application of a Resource Allocation System in a Technology-based Public Organization, R & D Management, Vol. 6, No. 1, 1975, pp. 31-37.

91. Shumway, C. R. and R. J. McCracken, "Use of Scoring Models in Evaluating Research Programs," American Journal of Agricultural Economics, Vol. 56, No. 4, November 1975, pp. 716-718.

92. Simmonds, N. W., "Costs and Benefits of an Agricultural Research Institute, R & D Management, Vol. 5, No. 1, pp. 23-28.

93. Souder, W. E., "A Scoring Methodology for Assessing the Suitability of Management Science Models," Management Science, Vol. 6, No. 2, pp. 81-86.

94. Steck, R., "How Can Research on Research Contribute to a Better Management of University Research?", R & D Management, Vol. 6, No. 2, pp. 81-86.

95. Sukhatne, V. A., "The Utilization of High Yielding Rice and Wheat Varieties in India: An Economic Assessment," Unpublished Ph.D. dissertation, University of Chicago, 1976.

96. Technical Advisory Committee, CGIAR, "Priorities for International Support to Agricultural Research in Developing Countries," TAC Secretariat, Food and Agricultural Organization of the United Nations, Rome, 1976.

97. Theil, H., "On the Optimal Management of Research: A Mathematical Approach," presented at the Conference of the International Federation of Operations Research Societies, Oslo, Norway, July 1963.

98. Thompson, R. L., "The Metaproduction Function for Brazilian Agriculture: An Analysis of Productivity and Other Aspects of Agricultural Growth," Ph.D. dissertation, Department of Agricultural Economics, Purdue University, 1974.

99. Thompson, R. L. and G. E. Schuh, "Sources of Regional Differences in Productivity in Brazilian Agriculture," presented at conference on Growth, Productivity, and Equity Issues in Brazilian Agriculture, Ohio State Uniersity, January 13-15, 1975.

100. Tweeten, L. G. and F. K. Hines, "Contribution of Agricultural Productivity to National Economic Growth," Agricultural Science Review, Vol. III, No. 2, 1965, pp. 40-45.

101. U.S. Department of Agriculture, "Manual of Classification of Agricultural and Forestry Research," Science and Education Staff, USDA, Washington, D.D., June 1970.

102. U.S. Department of Agriculture, "A National Program of Research for Agriculture," Report of a Study Sponsored Jointly by the Association of State Universities and Land Grant Colleges and U.S. Department of Agriculture, U.S. Government Printing Office, Washington, D.C., October 1966.

103. Wallace, T. D. and D. M. Hoover, "Income Effects of Innovation: The Case of Labor in Agriculture," Journal of Farm Economics, Vol. 48, No. 2, May 1966, pp. 325-336.

104. Welch, F., "Education in Production," Journal of Political Economy, Vol. 78, January-February, 1970, pp. 35-39.

105. Wennergren, E. B. and M. D. Whitaker, "Social Return to U.S. Technical Assistance in Bolivian Agriculture: The Case of Sheeps and Wheat," American Journal of Agricultural Economics, August 1977, pp. 565-569.

106. Williams, D. J., "A Study of a Decision Model for R & D Project Selection, Operations Research Quarterly, Vol. 20, No. 3, 1969.

107. Williamson, Jr., J. C., "The Joint Department of Agriculture and State Experiment Stations Study of Research Needs," in Resource Allocation in Agricultural Research, edited by W. L. Fishel, University of Minnesota Press, Minneapolis, 1971.

108. Wise, W. W., "The Role of Cost-Benefit Analysis in Planning Agricultural R & D Programs," Research Policy, Vol. 4, pp. 246-261.

REFERENCES

This is not a full bibliography on the subject but only a list of some references of interest for evaluations of agricultural research programs by national leaders.

Arndt, Thomas M, Dana G. Dalrymple, and Vernon W. Ruttan, (eds.)
Resources Allocation and Productivity in National and International Research. Minneapolis: U. Minnesota Press, 1977.

Bernsten, R.H.
Evaluating the Impact of Farming Systems Research and Development Projects. A neglected issue.
Paper prepared at the Farming Systems Research Symposium, Kansas State University, November 21-23, 1982, revised (mimeo)

Biggs, Stephen D.
Monitoring and control in agricultural research systems: maize in Northern India.
Research Policy: 12(1): 37-59, 1983.

Casley, Dennis J., and Denis A. Lury
Monitoring and Evaluation of Agricultural and Rural Development Projects.
Baltimore, MD: The John Hopkins University Press, 1982

Cernea, Michael and Benjamin J. Tepping
A system for monitoring and evaluating agricultural extension projects.
W.B. staff working paper no. 272. Washington D.C.: The World Bank, 1977.

FAO
Monitoring Systems for agricultural and rural development projects. 2 Volumes. Rome, Italy: FAO Economic and Social Development Paper no. 12, 1983.

Fishel, Walter, L. (ed.)
Resource allocation in agricultural research
Minneapolis: University of Minnesota, Press, 1971.

IBRD, Economic Development Institute.
The Institute publishes a series of course notes on all aspects of development. A complete list is available on request from the World Bank. Central Monitoring and Evaluation, CN-318, November 1980, is reprinted here as chapter VII.

Minnesota Agricultural Experimental Station
Evaluation of Agricultural Research
(Proceedings of a 1980 Workshop). Publication no. 8, 1981.

Murphy, Josette and Leendert H. Sprey
Monitoring and evaluation of agricultural change. Wageningen, The Netherlands: International Institute for Land Reclamation and Improvement, 1982.

Norton, George, and Jeff Davis
Review of methods used to evaluate Returns to Agricultural Research. Minnesota, Staff Paper P79-16, 1979

Patton, Michael Quinn
Utilization - focused evaluation.
London: Sage Publications Ltd., 1978

Patton, Michael Quinn
Qualitative Evaluation methods.
London: Sage Publications Ltd., 1980

Schuh, G. Edward, and Helia Tollini
Costs and Benefits of Agricultural Research: the State of the Arts.
The World Bank, Staff Working Paper no. 360, 1979.

USAID
Manager's Guide to Data Collection.
Washington D.C.: USAID/PPC/E, November 1979.