

Report No. 13247

An Overview of Monitoring and Evaluation in the World Bank

June 30, 1994

Operations Evaluation Department

FOR OFFICIAL USE ONLY



OPERATIONS EVALUATION DEPARTMENT
13247

FILE COPY

Report No: 13247
Type: OER

Document of the World Bank

This document has a restricted distribution and may be used by recipients only in the performance of their official duties. Its contents may not otherwise be disclosed without World Bank authorization

ACRONYM/ABBREVIATION

ADP	-	Agricultural Development Project (Nigeria)
AGR	-	Agriculture and Rural Development Department (Bank)
AGRME	-	AGR to Monitoring and Evaluation Unit
APMEPU	-	Agricultural Projects Monitoring, Evaluation and Planning Unit
APMEU	-	Agricultural Projects Monitoring and Evaluation Unit
Bank	-	World Bank
BIAS	-	Business Innovation and Simplification
CG	-	Community group
CPM	-	Computerized Project Management
CVPU	-	Central Vice Presidential Unit
ECDP	-	Evaluation Capacity Development Program
EDI	-	Economic Development Institute
EM	-	Electronic Mail
ERR	-	Economic Rate of Return
FIRA	-	Trust Funds for Agriculture
GTZ	-	Deutsch Gesellschaft fuer Technische Zusammenarbeit
HDM	-	Highways Development and Management Model
ICR	-	Implementation Completion Report
IEC	-	International Economics Department
IFAD	-	International Fund for Agricultural Development
IIM	-	Indian Institute of Management, Lucknow
KfW	-	Kreditanstalt fuer Wiederaufbau
KPI	-	Key Performance Indicators
LDC	-	Land Development Corporation
M&E	-	Monitoring and Evaluation
MIS	-	Management Information System
MOR	-	Ministry of Railways (China)
NGO	-	Non-governmental organization
NORAD	-	Norwegian Agency for Development
O&M	-	Operations and Maintenance
OD	-	Operational Directive
OECD	-	Overseas Economic Cooperation Fund (Japanese)
OED	-	Operations Evaluation Department
OM	-	Operational Memorandum
OMS	-	Operational Manual Statements
OPN	-	Operations Policy Note
OPR	-	Operations Policy Department
PBM&E	-	Planning, Budgeting, Monitoring & Evaluation
PAR	-	Performance Audit Report
PCR	-	Project Completion Report
PD	-	Project Document
PE	-	Participatory Evaluation
PHN	-	Population, Health & Nutrition
PIB	-	Project Information Brief
PMTF	-	Portfolio Management Task Force
PRDPH	-	Policy Research Department - Poverty & Human Resource Div.
PRISM	-	Program Performance Information for Strategic Management
RORSU	-	Rural Operation Review and Support Unit
RSAC	-	Remote Sensing Application Center
SAPS	-	Special Assistance for Project Sustainability (OECD)
SAR	-	Staff Appraisal Report
SL	-	Structured learning
TWU	-	Transport, Water and Urban Development Department
UNCHS	-	United Nations Center for Human Settlements
UNESCO	-	United Nations Economic and Social Council
UNIDO	-	United Nations Industrial Development Organization
USAID	-	United States Agency for International Development
VHO	-	Vehicle Operating Costs
WID	-	Women in Development
WSS	-	Water supply/sanitation
ZOPP	-	Zielorientierte Projektplanung

THE WORLD BANK
Washington, D.C. 20433
U.S.A.

Office of Director-General
Operations Evaluation

June 30, 1994

MEMORANDUM TO THE EXECUTIVE DIRECTORS AND THE PRESIDENT

SUBJECT: AN OVERVIEW OF MONITORING AND EVALUATION
IN THE WORLD BANK

Attached is the report entitled An Overview of Monitoring and Evaluation in the World Bank, prepared by the Operations Evaluation Department.

The record of Monitoring and Evaluation (M&E) in the Bank has been disappointing. The report discusses M&E plans and activities during appraisal as well as implementation, and finds that neither has responded well to Bank directives. However, the report also notes an improving trend across all sectors and regions. Promising work is also underway on sector level performance indicators and the implementation completion reporting guidelines recently issued will enhance demand for M&E. In general, the Bank's strategic focus on implementation and development impact should strengthen this momentum. So should institutional support measures geared to evaluation capacity development in borrowing member countries.

The report makes recommendations to help establish M&E as an indispensable tool of good management and part and parcel of the learning culture.

A handwritten signature in black ink, consisting of a stylized 'A' with a horizontal line through it and a vertical line extending upwards from the right side.

Attachment

This document has a restricted distribution and may be used by recipients only in the performance of their official duties. Its contents may not otherwise be disclosed without World Bank authorization.

AN OVERVIEW OF MONITORING AND EVALUATION IN THE WORLD BANK

Table of Contents

	Page No.
Preface	i
Executive Summary	iii
I. Introduction	
A. Background	1
B. Definitions	2
C. Study Design	3
D. The Report	4
II. Development of M&E Policies and Programs	
A. History	5
B. The Portfolio Management Task Force Report and the Next Steps	9
III. Project M&E at Appraisal: Late Development	
A. General Remarks	13
B. Characteristics of the Old and New Projects	14
IV. Project M&E in Practice: Disappointing Results	
A. General Remarks	26
B. Agriculture	
(i) The Golden Years	30
(ii) The Sample	32
(iii) The Case Studies	34
C. Other Sectors	43
D. Experience of Three Donor Agencies	45
(i) IFAD	46
(ii) USAID	46
(iii) GTZ	48
V. Recent Work above the Project Level: Remarkable Progress	
A. Sector Indicators	50
B. Participatory Evaluation	52
C. Recasting M&E in the Bank: the Learning Culture	54
VI. Findings and Recommendations	
A. Findings	59
B. Recommendations	
(i) Promoting M&E in the Bank	64
(ii) Reinforcing Next Steps	66
(iii) Organizational Responsibilities	68
(iv) Some Technical M&E Issues	68
(v) Further Study	69

	Page No.
Annexes:	
1. OED's Sample of Old and New Projects	71
2. "Monitoring and Evaluation in Agriculture ...", Gilroy Coleman, 1982	77
3. "India: Performance of State Monitoring and Evaluation Units Under the Training and Visit System of Agricultural Extension" Krishna, Jai and Raheja, S.K. April 1994	91
Boxes:	
1. Indicators for Managing Pavements	18
2. India: Uttar Pradesh Sodic Lands Reclamation Project	20
3. Yemen: Education Sector Investment Project	21
4. China: Sixth Railway Project	22
5. Brazil: Second & Third Northeast Basic Education Projects	23
6. Egypt: Matruh Resource Management Project	24
7. M&E in the Brazil Northwest Region Development Program	31
8. Data Collection and Use in China	33
9. M&E in the Indian T&V Agricultural Extension Program	35
10. M&E in the Brazil Northeast Rural Development Program	38
11. Examples of Bank Staff Skepticism about Substantial M&E	41
12. Ghana: Monitoring of Education Services	44
13. Kenya: Secondary Towns Project	45
14. Housing Indicators Program: South Africa	51
15. Participatory Evaluation in Water and Sanitation	53
16. Structured Learning in Water and Sanitation	57
17. Monitoring Project Environmental Impacts	58
Figure in Text:	
3.1 Quality of M&E in Old Projects: Ex-ante vs Ex-post	28
Tables in Text:	
3.1 M&E Content of Appraisal Reports: By Sectors	14
3.2 M&E Content of Appraisal Reports: By Regions	15
3.3 KPI Content of Appraisal Reports	16
4.1 Quality of M&E in Practice	26
4.2 Project Ratings during Supervision and at Completion vs. OED's Ratings of M&E	29
Tables in Annex 1:	
1. Distribution of Projects in the Desk Review	71
2. Completed (Old) Projects Included in the Desk Review	72
3. Recently Appraised (New) Projects Included in the Desk Review	74
Working Papers:	
India: Performance of State Monitoring and Evaluation Units under the Training and Visit System of Agricultural Extension (by CARDS - see Box 9).	
Case Studies of High-Profile M&E Programs in Agriculture (forthcoming, by OED).	

AN OVERVIEW OF MONITORING AND EVALUATION IN THE WORLD BANK

PREFACE

This study concentrates on the twenty-year history of World Bank involvement with the establishment and use of project monitoring and evaluation systems. The search for appropriate indicators and the monitoring of project performance covers a longer period. But it was only in the mid-1970s, when Bank priorities shifted to poverty reduction and rural development, that the collection of activities known as M&E received formal organizational recognition. The Overview examines monitoring and evaluation practices across sectors, including infrastructure sectors where the term "M&E" is not even used.

The Overview is based on inspection of Staff Appraisal Reports, Project Completion Reports and Performance Audit Reports covering two periods of activity since about 1980. Because interviews brought less new information bearing on M&E than expected, a group of case studies in the agriculture and rural development sectors was developed.

The Portfolio Management Task Force report and the Next Steps Program demanded that a study of monitoring and evaluation practices be current. Accordingly, the Overview focusses as much on present operational practice as on the record of past practice. Better appraisal designs for indicators and M&E systems are beginning to appear. There is evidence of the spread of a "learning culture".

Monitoring and evaluation are valued highly by other development agencies. The experiences of some of them are reported summarily, and their comments have been incorporated. It is planned to invite them and representatives of the recipient countries to join the Bank in a seminar at which the validity and implications of the report's findings can be further explored.

The study was undertaken by OED staff and international consultants. The Task Manager was Mr. Edward B. Rice, who prepared the report. Mr. McDonald P. Benjamin, Jr. provided research support. Ms. Silvana Valle and Ms. Isabel Alegre provided administrative assistance. The report has also benefitted from comments and reviews by present and past Bank operational staff and managers.

AN OVERVIEW OF MONITORING AND EVALUATION IN THE WORLD BANK

EXECUTIVE SUMMARY

Introduction

1. It has been Bank policy since the mid 1970s to promote monitoring and evaluation (M&E) as instruments of project implementation. The M&E tradition originated in agriculture and rural development, spreading gradually to other sectors with a social dimension. Many of the "hard" sectors do not use the term M&E, and concentrate on the monitoring dimension. A related operational requirement is the selection at the appraisal stage of "key performance indicators" (KPI), critical variables that can be used to measure implementation progress and development impact. Both M&E and the KPI tool were given prominence in 1992 by the report of the Portfolio Management Task Force (PMTF). Despite the policy support, these guidelines and directives were ignored or given inadequate attention at appraisal and in practice, especially in the social sectors where good M&E was considered most important. While recognizing some impressive exceptions, the overall results of the 20-year M&E initiative have been disappointing.

2. The situation has begun to change. Action to incorporate viable KPI and other M&E elements into project design has accelerated. The PMTF boosted a trend that was already underway. Other trends in the Bank's management agenda—in quality, participation, and the learning culture—suggest that the increase in interest in M&E may be part of a broader shift in Bank behavior accompanying the reorientation of the portfolio to equitable and sustainable

development. Nevertheless, the history of M&E in the Bank is characterized by non-compliance. The objective of the report is to determine why M&E is not in place after 20 years and what should be done.

3. The Overview is based on a desk review of selected project documents, subsequent discussions with regional and central staff, and field visits in several countries. A systematic study of "old" and "new" projects provided the basis for analysis. Samples were drawn from the two sets. The size of the samples was based on the prominence of M&E in the sector. Extra weight was given to those sectors where prior expectations indicated a prevalence of M and especially E activities. Thus, for example, about 20 old and 20 new agricultural projects were reviewed, 10 old and 10 new transport projects, etc. The totals were 89 old and 83 new, for a total of 172 projects. After the desk review, the study team of two OED staff interviewed selectively from among project officers concerned with the last years of the old projects and appraisal of the new, plus a smaller number of division chiefs and sector subject matter specialists.

4. The desk study and the interviews were not as strong a base for drawing lessons as had been expected. Bank staff have usually paid little attention to M&E activities, and the record of intentions and achievements as embodied in the SARs, PCR, and PAR is sparse. The interviews did not help. Staff that had not paid attention to M&E during the life of the project could not remember much about what they had ignored earlier.

5. To strengthen the Overview, OED extended the research in two directions. First, a number of other projects with high profile M&E programs, outside the samples but identified during the interviews, were brought into the desk review. Second, OED visited five countries to develop case studies of traditional agricultural programs recognized to have high profile M&E components, in India, Pakistan, Indonesia, Brazil and Mexico, and added consultant work from Nigeria and Malawi. OED feels that the sampling and interviewing—buttressed by the case studies—was adequate to assess the overall performance and trends in M&E, and to pick up for review as well many of the other vigorous M&E designs—outside the sample—that are now being put into play.

Development of M&E Policies and Programs

6. The report does not present in any depth the case for monitoring and evaluation. The principles are self-evident and secure, a position reaffirmed by PMTF. Better indicators strengthen monitoring, better monitoring lets management at both the project and financiers' levels identify delays and bottlenecks, and ongoing and ex-post evaluations find solutions to those problems and evidence of outputs and longer-term impacts. Also, M&E systems offer another opening to fix management's attention on social, institutional and participatory objectives. Doubts about the utility of M&E emerge from poor execution, not from principles. The Bank's directives have tried to put those principles into practice.

7. KPI have been dealt with in the directives on supervision. They were first proposed in the 1974 directive on Project Supervision and made mandatory by Operational Manual Statement (OMS) 3.50 in 1979. M&E has been dealt with in a separate set of instructions. In 1977 the Bank issued the first directive: OMS 3.55 recommended (but did not mandate) that all projects include

some form of both monitoring and evaluation. The Agriculture and Rural Development Department (AGR) confirmed its M&E leadership in 1974 by establishing a Monitoring Unit, renamed the Monitoring and Evaluation Unit (AGRME) in 1979. It was the first and only Bank unit ever formed with a mandate to support M&E. It was responsible for producing global KPI on rural poverty. But it provided support as well to regional staff in institution building: for creating capacity in project agencies for monitoring performance and measuring project impact. By the early 1980s, M&E was becoming part of standard operational practice in agriculture. But it did not yet have a strong hold anywhere else in the Bank and even in agriculture was about to come under attack. In 1983 the Bank initiated an exercise aimed at reducing excessive paperwork in its lending operations. For reasons explained in the text, the KPI annex was considered dispensable. OMS 3.50 was formally revised in 1985, at which time the requirement for a KPI annex was deleted. Part of the apparent inconsistency was due to the nominal separation of KPI from M&E: each was treated in a separate series of directives with no organic link and only casual cross reference. In 1985, AGRME set up a Working Group on M&E with active participation by supporters from other sectors. However, in the 1987 reorganization AGRME was abolished and a residual core staff was transferred to the new Operations Policy Department with a different responsibility. The connection with project M&E as well as agriculture was severed. The Working Group was disbanded in 1988.

8. But the fortunes of M&E were about to reverse again. In 1989, the two OMSs were reissued as Operational Directives. OD 13.05 mandated that KPI be reintroduced to all supervision reports; OD 10.70 mandated that M&E systems be incorporated in all future investment projects. The two activities were still separated, but they were now expected to be promoted Bank-wide. Enforcement was

again lax, such that many if not most projects continued to be processed through the Board with minimal or no attention to either indicators or the M&E system that was necessary to produce them. Nevertheless, a clear policy position on KPI and M&E had coalesced.

Project M&E at Appraisal

9. This chapter discusses the characteristics of M&E appraisal design in the sets of old and new projects. Two qualifications are in order. First, the SAR does not always reflect accurately the M&E/KPI content of the project. In the new set, a few projects have appraisal reports which are generally "quiet" on M&E/KPI, but subsequent interviews revealed an incipient M&E/KPI plan had already been discussed and sometimes been described in background and other working papers. Second, in the infrastructure sectors monitoring and indicators are often part of the project routine: task managers expect them to be included in the quarterly progress reports or demand they be made available to Bank field supervision. Many of these task managers ignored or were unaware of the Bank's directives to bring the indicators forward to the SAR. That behavior has adjusted radically to the post-PMTF era.

10. The first observation is the low frequency of significant M&E activity planned for the projects of either set. The 1989 OD called for effective M&E in all projects, but this mandate has been respected in less than half the projects where strong M&E should have been installed. Second, sectors with higher M&E activity can be distinguished. Agriculture was the leader in the early 1980s, but that position is now disputed. Education, PHN and water supply/sanitation have advanced relative to agriculture. On the other hand, the infrastructure sectors lag well behind these leaders, although there are some interesting exceptions at the project level. Third, in terms of the regional breakdown,

sharp distinctions in the old set have almost disappeared in the new. Africa was well behind in the old set—almost no projects with high M&E content, South Asia showed an even spread from no, low, to high content, and the other regions were positioned in between these two. For the new set, with one exception, all regions show a better profile.

11. With respect to the use of indicators—distinguishing that as one component of effective M&E—the rate was again well below a level reflecting satisfactory compliance with the OD. In agriculture, for example, only 1 of the 21 old SARs had a KPI table in its text or annexes. That statement refers narrowly to tables with indicators identified as such—a tool for guiding management and supervision in reviewing performance. Indicators were used more frequently in the old projects for the other social program sectors. For the hard sectors, for example the utilities and other infrastructure sectors, the incidence varies according to the nature of the product targeted by the project. For some sub-sectors, there has been substantial progress in the sophistication of the tools of measurement. Ports, telecommunications and highways and pavements are some good performers mentioned in the text.

12. The most frequently reported indicators are those demonstrating the state of financial health of the project agencies. These often dominate the SAR indicator lists. Furthermore, for the industry and utility sectors, financial KPI are considered by Bank professionals to be the essential test of project success. Since financial indicators are derived directly from routine MIS financial data, the other institutional requirements of M&E disappear. This position has been challenged on at least two counts. First, it is weakened by the common practice of public subsidization of consumer utility tariffs. Subsidies imply distortions in consumer uses and satisfactions. Second, the position is weakened wherever there are important externalities. Never-

theless, neither of these qualifications suggests that the project authorities themselves should be responsible for measuring economic impact. That is not what port and telecom authorities do.

13. Despite low rates of compliance, the tables in the text clearly show an increase in M&E content in the SARs in the approximately ten-year period between approval of the two sets of old and new projects. The percentage of SAR's with substantial M&E moves up from 10 percent to 36 percent, and those with substantial indicator content move from 12 percent also to 36 percent. The most important observation from the new set of SARs is that the ones with the highest M&E and KPI content are the haphazard product of spontaneous action by isolated individuals with a personal predisposition to put the tools of monitoring and evaluation to work. Division chiefs appear to have little to do with the *emergence* of this disparate group, though they *encourage* the innovators to continue. Small fires are burning in all of the social program sectors, and in some of the hard sectors as well.

14. Whether the more elaborate M&E operations planned in the new set will succeed is uncertain. Few of the SARs give any sense of government "ownership" of these components. Subsequent interviews revealed a high level of participation by government and beneficiaries in most—but not all—of the projects, and these assertions by the staff about country ownership are convincing. Nevertheless, the agreements are untested in practice, and *that* story, summarized in the next paragraph, shows very low levels of ownership during the 1980s.

Project M&E in Practice

15. After studying appraisal designs, OED looked for results: the performance of M&E programs in the old set of projects. Unfortunately, the institutional memory

problem was severe in interviews on the old set. The overall pattern was clear nevertheless: the Bank's record on the implementation of M&E is worse than the unsatisfactory performance already established at appraisal. OED's desk review shows that PCRs rated performance at or slightly better than appraisal content. OED's case studies of the seven older, continuing agricultural programs with substantial M&E components suggest that few of the improvements were sustained. Here, the SARs featured M&E, but the results measured in terms of data quality, useful conclusions, and management impact were all disappointing. That profile is repeated across all sectors where there was active M&E. There are exceptions at the project level, of course, suggesting that the problem with M&E in practice lies more with resources, incentives and implementation skills rather than errors of design. There are exceptions also from the period in the late 1970s and early 1980s *preceding* OED's old sample, when AGRME staff and their consultants (and a few practitioners in other sectors) brought their enthusiasm for M&E to the rural development and irrigation portfolios and demonstrated what good M&E can do. The study was looking at a later set, when direct Bank involvement had begun to fade. The Bank's disappointment is consistent with the academic literature that has emerged in recent years discussing the problems of M&E. It is also consistent with the experience at IFAD, the other development agency with a long and intense involvement in trying to establish M&E services at the project level.

16. Using the case studies as a base, OED tried to discern the factors which seemed best to explain poor performance. Three stood out: (1) the lack of a sense of ownership of the M&E program on the part of government, and, in particular, of project management; (2) a lack in continuity of attention by the Bank—in all cases there was no sustained Bank support for M&E after appraisal and the initial start-up period. These two factors—both

related to ownership—reflect a low demand for M&E products; and (3) deficiency in all cases as well in the appointment and retention of qualified nationals to the M&E unit. IFAD, in communication with OED on a draft of this report, emphasizes also the lack of incentives for all actors—aid agency operators, project managements, and field staffs—to concern themselves with errors and lessons and expose themselves to the consequences of reporting generally unimpressive project results. This reinforces the lack of ownership.

Recent Work above the Project Level

17. The report discusses some remarkable progress in activities associated with project M&E. Recognizing that "good projects depend on good sectors", the PMTF Report calls for the establishment of KPI lists at both the project and sector levels. Most of the early work for the ongoing Next Steps indicator exercise dealt with projects. But some good work has also been submitted at the sector level. The report gives examples of the conceptual treatment of diagnostic and enabling indicators proposed (and already in use) in industry/mining and housing. The report applauds as well a paper on indicators for the super-sector of poverty programs, another piece of work brought forward for Next Steps by the Operations Policy Department. Participatory evaluation in Bank projects—the process of including beneficiaries in the M&E operation—is another emerging theme that has prompted some good reporting. Several examples from the water and sanitation sector are given high marks. Finally and most encouraging, the report describes evidence of a gradual shift in Bank behavior towards a "learning culture". Here the M&E activities are internalized by Bank staff, monitoring and feeding back lessons of experience becomes the routine, and the phrase M&E drops out of the discussion. One of the central departments is now popularizing the concept of "structured learning", which reflects this new approach to managing projects. Structured learning and the learning

culture are alien to most Bank staff, but the expectation is that those ideas will spread and pull effective M&E along with them.

Findings

18. The major observations are brought together in the final chapter. Low compliance and low ownership (by the Bank as well as governments) heads the list. OED found broad agreement in the Bank on these points: except among the few veteran practitioners, the reputation of M&E is dismal. There are some encouraging exceptions, emerging from all sectors and regions, enough to signal an improving trend. It started before PMTF and Next Steps, and these will now push it faster. The threat to these rising expectations is the lack of any institutional capability—in the Bank, in the countries, and on the projects—to support the trend towards better M&E with professional advice. Other popular movements in the Bank—gender, environment, participation—have institutional and intellectual backup. M&E has no comparable support, though it (and its associated indicators) is skill intensive and anything but easy to implement effectively. This means that task managers who have to put M&E into appraisal design have nowhere they can turn to for guidance or training. As IFAD points out, the incentives for proper monitoring and evaluation have up to now also been missing at both the Bank and country levels.

Recommendations

19. The recommendations are grouped into four sections. First are those related to special actions to promote M&E in the Bank:

- a) create a capacity in the CVPUs and Regions to support the development of M&E systems;
- b) establish an M&E training program;
- c) recruit individuals with M&E, survey and statistical skills;
- d) network the M&E practitioners;
- e) focus on certain sectors; and

- f) create a temporary position at the center, to help launch these actions.

The second group reinforces measures already proposed by PMTF and Next Steps:

- (a) strengthen the role in M&E of the Project Advisor;
- (b) continue work on the policy and enabling indicators;
- (c) tie the indicators to the planning process, as in logframe;
- (d) strengthen ICR (PCR) reporting on M&E;
- (e) expand ECDP support for project M&E;
- (f) support post-completion monitoring;
- (g) integrate the ODs on indicators and M&E; and
- (h) promote participatory evaluation, judiciously.

One recommendation covers the institutional assignments:

- (a) the Regions should establish M&E positions, the CVPU should provide technical advice, OED should stand in support, PRD and other units should expand the number of "evaluative research" studies, and the Training Division with EDI support should prepare to move into M&E in strength.

Finally, three technical issues are addressed:

- (a) keep indicators bound in the context of M&E capacity;
- (b) put indicators to work – in projects where they can be precisely and collaboratively dimensioned – by conditioning and phasing disbursements; and
- (c) simplify or shift responsibilities for formal evaluations.

AN OVERVIEW OF MONITORING AND EVALUATION IN THE WORLD BANK

I. INTRODUCTION

A. Background

1.1 It has been Bank policy since the late 1970s to promote monitoring and evaluation (M&E) as instruments of project implementation. Operational Directives now require the use of M&E systems. Much of the M&E tradition originated in agriculture and rural development, spreading gradually to education, health, urban development and other sectors with a social dimension.

1.2 A related operational requirement is the selection at the appraisal stage of "key performance indicators" (KPI), critical variables that can be used to measure implementation progress and development impact. The KPI tool was given prominence in 1992 by the report of the Portfolio Management Task Force (PMTF).¹ Indicators are an important component of M&E. In the infrastructure and industrial sectors, key performance indicators (KPI) have been the primary design factors for M&E. The term M&E is hardly used. KPI have been oriented toward meeting the needs both of the Bank for aggregate performance data and of project management for improved implementation - with relative priorities shifting over time.

1.3 Despite the policy support, these guidelines and directives have been ignored or given inadequate attention. A common response was to prescribe a short list of indicators at appraisal without adequate institutional capacity at the project level to provide any but the easy information on inputs. In some instances large amounts of field survey data were assembled without the analytical capability to properly interpret them and apply the lessons. In short, and while recognizing some impressive exceptions, the overall results of the 20-year M&E initiative have been disappointing.

1.4 The situation has begun to change. Action to incorporate viable KPI and other M&E elements into project design has accelerated. The PMTF boosted a trend that was already underway. Other trends in the Bank's management agenda - in quality, participation, and the learning culture - suggest that the increase in interest in M&E may be part of a broader shift in Bank behavior accompanying the reorientation of the portfolio to equitable and sustainable development.

1.5 The main objective of the Overview is to review historical developments and present practices in M&E in the different sectors, to identify the factors that have limited or encouraged progress to date, and to recommend measures to make M&E effective. The case for effective M&E is widely accepted and discussed at length in PMTF's report. Given the focus on results proposed by the President and endorsed by the Board under the Next Steps program of action,² the role of

1. "Effective Implementation: Key to Development Impact." September 22, 1992.

2. "Portfolio Management: Next Steps, A Program of Actions." OPR. July 22, 1993.

M&E in Bank activities is certain to expand. The objective of the report is to determine why M&E is not in place after 20 years and what should be done.

1.6 This is the sixth report on project monitoring and evaluation prepared by OED. The first was dated 1977, and the most recent 1985.³ Four of those have dealt with agriculture, irrigation and rural development. One dealt with education. This is the first review to look over all sectors, and, with one partial exception, is the first such global study of M&E in the Bank. In 1985, in response to OED's agricultural M&E review, the Operations Policy Staff briefly discussed progress in some other sectors. The response was coordinated by the central agricultural M&E Unit which was established under another name in 1974 and has since been dissolved (paras 2.6, 2.10). In this report, OED regains a wider perspective.

B. Definitions

1.7 The report discusses Bank experience over the spectrum of monitoring and evaluation activities commonly incorporated in the term M&E. Monitoring refers to observations of physical and financial progress and external conditions which affect progress. It includes also measures of the first – and higher – order outputs of the projects. Evaluation refers to the analysis of project-related data to find causes, relationships and lessons. An appraisal design which calls upon management and the Bank to carry out only mid-term and completion reviews is not included in the definition of an effective M&E operation, since the institutional element and continuous feedback process is largely ignored. The border between monitoring and evaluation is not sharp. For what other reports call "ongoing" or "concurrent" evaluations the Bank's M&E guidelines⁴ use the phrase "diagnostic studies" and include them on the monitoring side. These guidelines restrict the term "evaluation" to mid-term and ex-post analysis, and thereby open the way to separating that function from monitoring conceptually as well as organizationally. The issues of whether to separate and where to locate the different M and E activities are important to this report, and are picked up again in paras 6.9ff. But the treatment is incomplete. A separate study would be needed to provide a rigorous assessment of the relative advantages of alternative options in varying circumstances.

1.8 The report overlaps the much broader field of "management information systems" (MIS). During the desk review, notes were taken of references to Bank and project support to improving MIS systems in the executing agencies. The term MIS refers here to all project-based information systems that control the quantification, aggregation, communication, storage and uses of data generated - from project activities, from the agency's full line of operations, and from its administrative actions. As the study progressed the emphasis narrowed - to the uses of MIS records to support purposeful monitoring and evaluation. There are interesting issues related to the entire field of MIS which would also benefit from further evaluation.

-
3. "Built-in Project Monitoring and Evaluation: First Review (Agric and RD)." October 14, 1977. Report #1758;
 "Built In Project Monitoring and Evaluation: A Second Review (Education)." November 2, 1979. Report #2724;
 "Built-in Project Monitoring and Evaluation: Third Review (Irrigation)." February 2, 1981. Report #3320;
 "Brazil - The World Bank - Built-in Project Monitoring and Evaluation: Rural Development in NE Brazil." May 1984. Report #5078;
 "Built-in Project Monitoring and Evaluation: an Overview." June 28, 1985. Report #5781.
 4. "Project Monitoring and Evaluation In Agriculture." Casley, Dennis J. and Kumar, Krishna. Johns Hopkins Press. 1987.

C. Study Design

1.9 The report discusses Bank experience over the spectrum of monitoring and evaluation activities commonly incorporated in the term M&E. It is based on a desk review of selected project documents, subsequent discussions with regional and central staff, and field visits in five countries.

1.10 A systematic study of "old" and "new" projects provided the basis for analysis. Samples were drawn from the two sets. The old set was defined as those projects with the most recently dated PCRs or PARs - up to a June 30, 1993 cutoff. The new set was defined as those projects most recently approved by the Board, up to the same June date. Exceptions were made - for example a few categories of projects were excluded, including all the adjustment operations,⁵ and a few selected projects were dropped if the country was already represented in the set for that sector. By and large the "June cutoff" rules were respected. Samples were drawn from all except the excluded categories. The size of the samples was based on the prominence of M&E in the sector. Extra weight was given to those sectors where prior expectations indicated a prevalence of M and especially E activities. Thus, for example, about 20 old and 20 new agricultural projects were reviewed, 10 old and 10 new transport projects (of each group, half were roads and half were railroads and ports), and 3 old and 3 new oil and gas projects. The totals were 89 old and 83 new, for a total of 172. In each case OED restricted its documentary sources to the Staff Appraisal Reports (SARs) and, for the old projects, Project Completion Reports (PCRs) and Performance Audit Reports (PARs, where they existed). Annex 1, Table 1 shows the breakdown and sectoral and regional subtotals. Annex 1 also lists the projects in the sample. After the desk review, the study team of two OED staff interviewed selectively from among project officers concerned with the last years of the old projects and appraisal of the new, plus a smaller number of division chiefs and sector subject matter specialists.

1.11 The desk study and the interviews were not as strong a base for drawing lessons as had been expected. The main problem is suggested in para 1.3: Bank staff have usually paid little attention to M&E activities, and the record of intentions and achievements as embodied in the SARs, PCRs, and PARs is sparse. The interviews could not adequately compensate for that deficit in the written record. Staff that had not paid attention to M&E during the life of the project could not remember much about what they had ignored earlier. Larger samples and wider interviewing would not correct for that problem. Examination of supervision, consultant and project progress reports would have helped fill the gap, at the cost of a substantial extension of the time allocated with uncertain incremental benefits.⁶

1.12 To strengthen the Overview, OED extended the research in two directions. First, a number of other projects with high profile M&E programs, outside the samples but identified during the interviews, were brought into the desk review. Second, OED visited five countries to develop case studies of traditional agricultural programs recognized to have high-profile M&E components - extension in India, irrigation and drainage in Pakistan, irrigation in Indonesia, rural development in Northeast Brazil, and credit in Mexico. In addition, consultants were used to update the record of

5. Including emergency programs, technical assistance and DFIs, along with the adjustment operations. Structural adjustment projects were excluded on the assumption that any KPI agreed as conditionality on tranche releases received better treatment than they did for investment projects.

6. The project files, including those three sets of reports, always add a rich background for understanding project events, but are not a conclusive source for comparative analysis.

two of the largest African rural development M&E programs - in Malawi and Nigeria. Only one of these seven programs had been captured in the sampling procedure. In two of the five countries visited by OED, brief inspections were also made of new operations with high profile M&E proposals in education (Northeast Brazil) and water supply and sanitation (Indonesia and Northeast Brazil).

1.13 The research was relatively rapid, by OED standards for its "studies", and the report does not pretend to have canvassed the majority of old and new activities in each sector. Nevertheless, OED feels that the sampling and interviewing - buttressed by the case studies - was adequate to assess the overall performance and trends in M&E, and to pick up for review as well many of the other vigorous M&E designs - outside the sample - that are now being put into play.

D. The Report

1.14 The report has six chapters. Chapter II describes the historical development of M&E policies and institutional responses within the Bank. The next two chapters discuss the M&E content at *appraisal* - comparing the old and new sets (Chapter III) - and in *practice* - looking at the actual record of M&E across the old set and the case studies (Chapter IV). Chapter V discusses several new initiatives in the M&E arena, or associated with M&E, which suggest that the upward trend represents a secular and positive shift in the Bank's M&E behavior. It reviews recent work with KPI at the sector level, and then discusses the elements of a "learning culture" that are spreading within the Bank. The final chapter presents the findings and recommendations.

II. DEVELOPMENT OF M&E POLICIES AND PROGRAMS

A. History

2.1 Monitoring and evaluation activities have been inserted in investment projects, with varying intensity and frequency, since the inception of Bank lending. In some sectors the Bank has routinely monitored implementation and reported against a set of progress indicators. Key data, for example turnaround-time of ships at berth in port development projects, can be such an integral part of project rationale that its neglect at appraisal and supervision would not have been tolerated. Whether these KPI were also arrayed in an annex with that name in the SAR was another matter: their absence there did not imply they were ignored in implementation. As the Bank in the early 1970s shifted priorities toward reducing rural poverty, the absence of a similar set of conventional indicators for rural development prompted the Bank to consider mandatory indicators in all sectors.

2.2 The history of the directives dealing with KPI and M&E - the Operational Memorandum (OM), the Central Projects Memorandum (CPM), the Operational Manual Statements (OMS) which replaced the OM in 1974, the Operations Policy Notes (OPN) which supported the OMS, and the Operational Directives (OD) which replaced the OMS in 1989 - gives a rough approximation of the trends in Bank policy.

2.3 KPI have been dealt with in the directives on supervision. They were first proposed in the 1974 OM on Project Supervision (the "Project Supervision Handbook").⁷ A set of indicators for a representative railroad project was included for illustrative purposes as an annex.

"Supervision of the borrowers' operation, as distinct from progress in constructing the project, may be facilitated by obtaining reports on a few key performance indicators which are the best evidence of the degree of success of the project or the resolution of problems identified at appraisal. Such indicators should be tailored to each specific situation. Typical indicators for several sectors and an illustration of how operational targets can be made a part of a loan agreement for a railway project are shown in Annex A" (page 9).

The 1979 OMS 3.50 on Project Supervision dropped the optional feature and mandated that all reports include a key indicator table among the annexes:

"The indicators selected should, of course, be the ones most appropriate for the specific project but it is expected that they will be similar within sub-sectors. If a monitoring and evaluation system has been set up, the indicators will be a sample of those used for monitoring" (Attachment B, page 1).

2.4 M&E has been dealt with in a separate set of instructions. In 1977 the Bank issued the first directive on Project Monitoring and Evaluation. OMS 3.55 recommended (but did not mandate) that all projects include some form of both monitoring and evaluation:

7. In 1974 also, the new Monitoring Unit in AGR (para 2.6) introduced a Project Information Brief (PIB) that put quantitative supervision data into a standard format for purposes of aggregation and global monitoring of the agricultural part of the rural development portfolio.

"Monitoring and evaluation are part of the efforts that are being made by borrowing countries and the Bank to implement projects more efficiently, to review their progress more closely and to ensure more effectively that the projects meet their development objectives...The purpose of this Statement is to clarify a number of concepts related to monitoring and evaluation and to provide general guidelines about their design and implementation" (page 1).

2.5 During the last part of the 1970s the Bank extended M&E throughout the rural development portfolio. The OMS on mandatory KPI and the OMS recommending M&E⁸ supported the Bank in its attempt to get measures of progress in the poverty-oriented portfolios that it could provide to the Board and external audiences. M&E was also spreading, slowly, in the new education, health, and urban portfolios, as those sector programs expanded. The directives had a smaller incremental impact on the infrastructure sectors, where, as stated above, monitoring of certain key indicators was common anyway - whether or not they were made explicit in the SAR - and staff considered ex-post evaluations unnecessary.

2.6 The Agriculture and Rural Development Department (AGR) confirmed its M&E leadership in 1974 by establishing a Monitoring Unit, elevated two years later to divisional level as the Rural Operations Review and Support Unit (RORSU). It was renamed the Monitoring and Evaluation Unit (AGRME) in 1979. It was the first and only Bank unit ever formed with a mandate to support M&E. It was responsible for producing global KPI on rural poverty. But it provided support as well to regional staff in institution building: for creating capacity in project agencies for monitoring performance and measuring project impact. From 1977 on the unit had a professional staff of about six. They undertook frequent trips to help set up and steer the new M&E activities, especially in rural development and irrigation programs. In 1979 AGRME also helped organize three overseas seminars - in Kenya, Costa Rica and Malaysia. It continued to sponsor workshops intermittently over the next seven years - for regional gatherings, for individual countries and for subsector activities (for example M&E for Extension Projects in the West African coastal zone).

2.7 OMS 3.55 on M&E proposed that manuals be prepared on monitoring and evaluation specific to each sector, with associated lists of representative indicators. AGRME staff had drafted these instructions. After two years of preparation, it issued in 1981 a softcover *Handbook* and companion *Guidelines* for M&E for agriculture.⁹ They were well received by other development agencies and borrowers as well as within the Bank. Further work led eventually to a series of two hardcover volumes, on project M&E and on data collection and analysis, sponsored by FAO, IFAD and the Bank.¹⁰ They are presently accepted as standard references for agriculture and rural development. The Water Supply and Urban Development Department followed in 1986 with a

8. Together with the PIB: see footnote 7.

9. "A Handbook on Monitoring and Evaluation of Agriculture and Rural Development Projects." Casley, D.J., and Lury, D.A. World Bank, November 1981.
 "Guidelines for the Design of Monitoring and Evaluation Systems for Agriculture and Rural Development Projects." World Bank, November 1981.

10. "Project Monitoring and Evaluation in Agriculture" (see footnote 4) and "The Collection, Analysis, and Use of Monitoring and Evaluation Data." Casley, Dennis J., and Kumar, Krishna. Johns Hopkins Press, 1988.

Handbook for Bank practitioners in urban development.¹¹ No other sector responded to the OMS proposal for sector-specific M&E guidelines. Indicator lists subsequently appeared for several sub-sectors - electric power, telecommunications, housing - but these latter-day exercises, aimed only at KPI, were prompted by perceived needs related to those sub-sectors and not by the earlier Bank-wide campaign. A coordinated effort to consolidate KPI across all sectors was not resumed until 1993, in response to the Next Steps program.

2.8 By the early 1980s, M&E was becoming part of standard operational practice in agriculture. But it did not yet have a strong hold anywhere else in the Bank and even in agriculture was about to come under attack from an unexpected direction. In 1983 the Bank initiated an exercise aimed at reducing excessive paperwork in its lending operations. Supervision, and in particular supervision reporting, came under scrutiny, and emphasis was given to streamlining the supervision report and dropping nonessential components and annexes. For reasons explained in the next paragraph, the KPI annex was considered dispensable. OMS 3.50 was formally revised in 1985, at which time the requirement for a KPI annex was deleted. The shift in the rules on this issue was recognized before the decision was formalized, so that in the years immediately before the 1985 OMS was issued, the practice of including KPI tables - in the sectors where it had begun to become common - had already started to lapse. This was true even in agriculture - despite the fact that at the same time the new M&E guidelines were being promoted. In fact the retreat from KPI in agriculture was encouraged by AGRME, which had begun to doubt the utility of sector specific indicators.

2.9 Part of the apparent inconsistency was due to the separation of KPI from M&E: each was treated in a separate series of directives with no organic link and only casual cross reference. Another problem was a sense among operationally-oriented professionals in the Bank that the 1974 requirement that all projects produce KPI, coupled to the 1977 recommendation that each sector produce indicative indicator tables for general use in projects throughout the sector, led to simplistic characterizations of inherently complex projects. The operational utility of the generalized indicator set was deemed not worth the effort of getting a group of professionals in each sector to agree on a common list. Thus, when the task force proposed to streamline the supervision report by deleting the KPI tables, there was no concerted move by M&E supporters to block the action.

2.10 Inconsistency in Bank actions on M&E and KPI appears again in the last part of the decade. In 1985, AGRME set up a Working Group on M&E with active participation by supporters from other sectors. In the 1987 reorganization, however, AGRME was abolished. Its chief and some other staff were transferred to the new OPR complex with a different responsibility, this time for monitoring the Bank's overall lending portfolio. The connection with project M&E as well as agriculture was severed. Technical assistance to the regions stopped, and most of the agricultural staff of AGRME were released. The Working Group was disbanded in 1988.

2.11 Before its removal from AGR, the AGRME team had begun the revision of OM 3.55 on M&E, basing the new statement on some of the principles set forth in its hardback volumes. Simultaneously, a reaction to the deletion of the KPI from supervision reports developed in the operational complex: task managers and division chiefs recognized that an important instrument for assessing progress had been lost. In 1989, the two OMSs were reissued as Operational Directives. OD 13.05 mandated that KPI be reintroduced to all supervision reports: "to display in tabular form

11. "Monitoring and Evaluating Urban Development Programs, A Handbook for Program managers and Researchers." Bamberger, Michael, and Hewitt, Eleanor. World Bank Technical Paper Number 53. 1986.

quantitative measures of project performance in critical areas" (Annex D4, page 1). OD 10.70 mandated that M&E systems be incorporated in all future investment projects.

"Plans for monitoring and evaluation are to be included in all Bank-funded projects, but their relative emphasis, scope and organization will vary, depending on the project and the implementing agency. These plans should be analyzed as early as possible during the project cycle, but no later than the appraisal stage, and all the arrangements must be agreed at negotiations..." (page 2).

2.12 The two activities were still separated, but they were now expected to be promoted Bank-wide. Enforcement was again lax, such that many if not most projects continued to be processed through the Board with minimal or no attention to either indicators or the M&E system that was necessary to produce them. Nevertheless, a clear policy position on KPI and M&E had coalesced. The policy was pushed along by growing concern over an apparent deterioration in the quality of the portfolio under implementation, the same concern that led to the appointment of the Portfolio Management Task Force in February, 1992. Monitoring of project performance moved to center stage.

2.13 The Economic Development Institute has had an occasional role in training for M&E. It was covered as a component in the standard course on the project cycle held in Washington in the 1970s and early 1980s. Subsequently, as EDI shifted its attention overseas, M&E was again featured as a component in its agricultural sector management courses. Prominent among these were the *Agricultural Management Training in Africa Program*, in collaboration with IFAD, aimed at projects' staff from individual countries. A second, highly successful program was the series of annual courses for the agricultural bank and ministries in China from 1982 to 1986. Nepal is a country which has also received special attention - in the late 1980s. In this case the training was aimed specifically at M&E. The target populations of these overseas courses were personnel from ministries, agencies and NGO's involved with central and project level management and M&E systems. Bank staff participated, though EDI never ran any of its courses specifically for that staff. However, there has been no EDI activity in M&E for the last three years.

2.14 Reference should be made also to a substantial textbook on M&E that has been under preparation in EDI for the last four years - *Social Program Evaluation in Developing Countries*. One of the two authors was also responsible for the M&E handbook on urban projects and the present exercise is an extension of that earlier work. It is a guide for policy makers, managers and practitioners, and offers extensive treatment of, among other subjects, both rapid and formal survey methodology. The term "social programs", a category which includes small farm agriculture, is a useful grouping and referred to elsewhere in this report.¹²

12. The domain of this EDI textbook is described in the following quote from the preface. "The term "social programs" refers here to the broad range of programs designed to improve the quality of life by improving the capacity of citizens to participate fully in social, economic, and political activities at the local or national levels. On the one hand, such programs may focus on improving physical well-being (health, nutrition); access to services (housing, water supply, local transportation); protecting vulnerable groups from the adverse consequences of economic reform and structural adjustment; or providing education, literacy, and employment and income-generating opportunities (vocational and technical training, credit, integrated rural development, small business development). On the other hand, they may focus directly on local empowerment and equity issues by strengthening community organizations, encouraging women to participate in development, or alleviating poverty."

2.15 A final point of historical relevance concerns not the field of M&E but of statistics in the Bank. The Bank never has supported in its institutional structure a department or division responsible for oversight on methodology and quality of statistics. There are two divisions in the International Economics Department (IEC) whose business is to collect and, for some functions, generate statistics. And statistical data bases and econometric analysis are in common use throughout the agency. But there is no institutional home for professional statisticians with broad experience in survey method and survey data quality, and no place for task managers to go to get advice and support on statistical technique (except the Africa Region: see below). At present there is a statistical advisor in IEC, whose jobs include liaison with other agencies, development (with those other agencies) of common methodologies for preparing national statistics, participation in designing research programs with statistical properties, and support to operational departments in improving national statistical capabilities at the borrower level. But he is not used as a resource person for the rest of the Bank for reviewing proposals for project level surveys and other statistical operations (it is difficult to imagine a single individual handling that job). He was not included in the network for the ongoing indicator activity under Next Steps. The only group of survey statisticians (five, in all) at present in the Bank is handling the residual core activities of the Social Dimensions of Adjustment Program (SDA) in the Africa Technical Department. The work of this group is evolving toward general technical assistance on statistical survey to the African Departments. But this latter function is unique in the Bank: there is no comparable Bank-wide team and none of the other Regions have set one up. The Bank has traditionally taken the view that its comparative advantage is not in statistical methods, and that it should leave that to the agencies which do claim competence, particularly the United Nations.

2.16 The point is made here to fill in the background to events in the field of M&E. For example, it is improbable that the indicator exercises of either the early 1980s or of Next Steps would have been carried out without the support of the statistical department, had it existed. This would automatically have brought together professional statisticians, with their sense of the quality and feasibility of data supporting the indicators, with the economists and sector specialists searching for KPI relevant to measure poverty, progress and other Bank targets. This consortium of talent would have substantially reduced the problems inherent in any drive to identify viable KPI. It is one thing to let other agencies take the lead in statistical sciences. It is another not to make use of those services when needed. The dangers inherent in the disconnect can be overcome, but only if the problem is well recognized. It is recognized, but maybe not well enough.

2.17 The similarity between the casual treatment of both statistical method and M&E is probably not accidental. High level commitment to one would imply commitment to the other. Also, if the Bank had had a strong statistical arm in the early 1970s, before the interest in poverty indicators and M&E in rural development accelerated, the form, targets, and professionalism of the Bank-wide M&E operation would have been grounded on surer footing. And the Bank probably would have given less emphasis to starting-up M&E survey work in ministries of agriculture and project management units - which had no prior experience with data collection - and more attention to building the competence of the existing national statistical services.

B. The Portfolio Management Task Force Report and the Next Steps

2.18 The report - *Effective Implementation: Key to Development Impact* - was issued in September 1992. The main text of 33 pages is supported by annexes, of which Annexes A, C and D are germane to this review. Annex C is the most important - "Towards a Results-Oriented Evaluation and Rating Methodology for Bank Supported Operations". The whole package is a major force for reform of

Bank practices both before and during implementation: to guarantee that attention remains fixed on the objectives of the project and on removing the constraints limiting expected progress towards those objectives. PMTF's terms of reference had been focused by "the belief that inadequate attention to monitoring and supervision of projects was a major factor behind the deteriorating performance of the Bank's portfolio".¹³ Its report strengthens the position of supervision vis-a-vis appraisal, and puts the quality of lending over quantity. It has been well received in the Bank, and will help hold both KPI and M&E at center stage.

2.19 The report, however, does not adequately address the relationship between the KPI and M&E. The main text and Annexes A and D refer to the requirements - of project management - for a proper monitoring system and reliable indicators. Annex C, which is the technical document supporting the proposed indicator system, says that indirectly:

"..the focus of this paper is on Bank actions and processes. This is not to preclude the involvement of borrowers. Quite the contrary. Building local capacity for evaluating investments - both in the context of public expenditure reviews and sector investment loans - is an ultimate goal... But before proceeding to that stage of the exercise, we need first a methodology appropriate to the times. Once broad consensus is reached within the Bank, we can proceed with dissemination" (Annex C, page 2).

2.20 But the spotlight of Annex C, and of the whole Report, is on developing the *Bank's system* for tracking progress towards the project's ultimate objectives. The importance of developing a monitoring system for *project management* - to track progress along the critical path of project execution, and to assess performance against a range of intermediate input and output targets - is acknowledged but given less attention. These priorities are partly explained - as stated in the last quote - in the interest of first positioning the Bank. Elsewhere in its report, PMTF calls for the regions to assume primary responsibility for the Evaluation Capacity Development Program (ECDP), taking over the ownership role from OED. ECDP is intended to strengthen the borrower's evaluation system as a whole. In principle, it should impact upon project M&E work, and, hence, on the indicators. But Bank-borrower activity to date on ECDP has concentrated on evaluation capacity at the national and ministerial levels. The linkage to project level M&E is not highlighted in the report either. Most professionals in the field of M&E are concerned that putting KPI above or before M&E gives the wrong signals, especially when the KPI are or appear to be Bank-oriented. They say that to defer country capacity building while emphasizing Bank indicators is a misguided strategy.

2.21 The PMTF report's proposals for KPI include two features that extend beyond the earlier guidelines on indicators. First, the KPI are not so much the descriptive indicators themselves, e.g. increased rates of employment for graduates of the improved vocational training program, but the benchmark values that are set for these indicators as annual and final targets for the operation, e.g. that average annual earnings will exceed by 40 percent those of a comparable group of graduates from traditional vocational courses. These numerical KPI can be treated as tools to structure subsequent discussion with government and project management. But they can be used also as benchmarks to identify weaker project components that need to be adjusted or dropped. The PMTF report, especially Annex C, conveys this aggressive role. The 1974 Handbook on Supervision, including OMS 3.50, mentioned such a use for KPI. OED's 1992 report on supervision also regretted

13. "Performance Indicators to Monitor Poverty Reduction". Carvalho, Soniya, and White, Howard. Second Draft, October 1993. Page v.

"the absence of automatic "flags", which would signal the need for some form of action".¹⁴ Proponents of aggressive KPI argue that is what the reform is all about: to give project management and the Bank tools for spotting and correcting non-performing components or if necessary cutting them out. The majority of Bank staff are reluctant to specify threshold values that can have an automatic and punitive impact. They feel the KPI are too unreliable a tool to trigger radical surgery. But as the Bank's experience and success with conditional indicators improves, in the subset of projects for which they are appropriate, and provided they are not used mechanistically, those doubts should gradually fade.¹⁵

2.22 The second, and more novel feature of PMTF's treatment of KPI is to call for those that can be tracked even after the project has closed and supervision has ended. Except when a project is succeeded by a follow-on operation, the concept of depending on government and the executing agency (and project management, if it still exists) to continue to report to the Bank on project impact indicators is not only novel but, in the opinion of most supervision staff, unrealistic under present conditions. PMTF was conscious of the difficulty in implementing its proposal, but was driven to it by clear evidence that in its absence the Bank would never really appreciate the long run effects of its lending and that supervision ratings up to project completion could be grossly misleading:

"This paper was also motivated by a related finding of the Wapenhans Report of a "worrying and growing" discrepancy between the generally favorable supervision ratings during implementation and the subsequent, less favorable, evaluation ratings at completion - satisfactorily implemented projects seemed to have failed upon completion. Accordingly, the Wapenhans Report underlined the need to monitor development impact during the operational phase of the project"¹⁶

2.23 In projects where the implementation phase overlaps the operational phase, impact monitoring can be built-in. In other projects, such as turnkey projects, monitoring of the impact of the operations supported by the project can be carried out only after "project completion". In these cases, the tracking routine will continue only if government sees some self-interest in carrying on with the data collection. Without government "ownership", the PMFT proposal would appear problematic. The task for the Bank is to encourage governments to recognize their interest, in common with the Bank, in monitoring activities during the operational stage. The Implementation Completion Report (ICR), which is replacing the Project Completion Report (PCR), is intended to do just that. So is ECDP.

2.24 Bank management responded formally to the PMTF Report in its document *Portfolio Management: Next Steps, a Program of Action*, issued in July 1993. Indicators, monitoring and evaluation are central to, but nevertheless only a part of, the PMFT recommendations. Next Steps also covers the field. But it pushes hard on the Bank's obligation to develop appropriate indicators and a tracking system locked onto those indicators for Bank supervision. It does not mention capacity building at the project level, or monitoring for project management, and in those respects reflects the

14. "Bank Experience in Project Supervision." April 30, 1992. Report #10606.

15. Structural adjustment operations were excluded from this review precisely because of the conditionality feature attached to their indicators (see footnote 5). By this omission, important lessons may have been missed on how to use aggressive indicators effectively.

16. Carvalho and White, op. cit., page v.

imbalance in the PMTF Report (however, it does embrace its recommendations on ECDP). This is a correctable fault, as suggested in the quote from Annex C above. But thus far Next Steps activity at the project level has concentrated exclusively on KPI.

2.25 The Bank placed responsibility for organizing and promoting the action program outlined in Next Steps with the Operations Policy Department (OPR). OPR instructed all sectoral departments, under the umbrellas of the three new Central Vice Presidential Units (CVPU), to prepare indicative sets of KPI appropriate for each sector in time to be submitted to all operational staff in FY94. An indicator task force was established, under the leadership of OPR, and subsector networks were set up in many of the departments. Progress accelerated in the early months of calendar 1994, and the overall exercise amounts to a substantial advance beyond the last indicator high-tide in the early 1980s. The skepticism over the usefulness of sector-specific KPI - so evident in the early 1980s - has diminished as the work progressed. As discussed below, some of the presentations already submitted are impressive contributions to the "science" of indicators. Most of the early Next Steps indicator work was concentrated on project-level indicators. But Next Steps call for indicator sets appropriate at both project *and* sector levels, and work is advancing on that larger dimension too. The work involves an increasing number of operational staff. This is a direct and positive outcome of the PMTF, and has already put the Bank well ahead of where it got to with KPI a decade earlier.

2.26 Except for ECDP, no corresponding work on M&E concepts, methods and country capacity is underway in the Bank. Some of the sector KPI submissions to OPR state clearly that these issues must be addressed. Other submissions are silent on the subject, and among these are a few which specify indicators that would appear to have no chance of ever being collected (that are worth anything!).

2.27 PMTF supports KPI as an essential tool for good planning as well as project management. The KPI requirement starts at project preparation, and forces corresponding improvements in defining objectives and establishing time-bound targets for reaching them. PMTF recommends that preparation employ the Logical Framework ("logframe"), a popular planning methodology whose tabular, hierarchical structure forces the planner's attention onto specifying targets, indicators, and assumptions about necessary and sufficient conditions for successful implementation. The quality of the project as appraised ("quality at entry") is improved to the extent the objectives - and measures to assess progress towards those objectives - can be made explicit early and determine the indicators and the monitoring design. Logframe does that. There are supporters and critics ("simplistic") of logframe in the Bank, and the PMTF's endorsement helps push the resistance aside. The report returns to this subject in Chapter IV.

2.28 The Africa Region is the most aggressive in applying logframe, by itself or as a component of a broader planning approach referred to as Computerized Project Management (CPM). This Region has distinguished itself in several other fronts - as it advances toward better portfolio management (paras 2.15, 5.11 and Table 3.2).

III. PROJECT M&E AT APPRAISAL: LATE DEVELOPMENT

A. General Remarks

3.1 This chapter discusses the characteristics of M&E appraisal design in the sets of old and new projects. The old set contains projects that were approved in the first half of the 1980s; the new set contains projects approved in 1993. Thus, the chapter compares the M&E content of two batches of SARs finalized about a decade apart.

3.2 Four qualifications are in order. First, as mentioned above, most of the "old" set was approved after a *decline* of KPI and M&E activity. Since the average age of the projects was about nine years, the June 1993 cutoff for PCRs and PARs meant that the majority were appraised after the high tide of activity in the late 1970s and early 1980s. Two previous Bank reports on M&E show higher percentages of agriculture projects with M&E: a 1976 "Progress Report" on M&E found that 30 of the 38 rural development projects (79 percent) approved in 1975 had M&E components;¹⁷ a later report from AGRME said that 89 percent of the same category of projects approved in the first years of the 1980s carried M&E components.¹⁸ In *quantitative* terms the M&E activity in that earlier period appears to have been running well ahead of the activity in *both* periods under inspection in this review.

3.3 Second, the "new" set of projects was prepared and largely appraised before the PMTF report was issued. PMTF deliberations took place while most of these projects were in appraisal status: the work was already in motion and was not significantly influenced by the report. For almost all of the new projects with high M&E content included in the review, the task managers claimed that their attention to M&E and KPI predated PMTF. The test of the report's influence will be on future projects. The PMTF impact is not under review here except as occasional evidence of early impact was brought up in a few interviews. OED might return to this issue with a comprehensive look at the report's influence on subsequent M&E design at appraisal.

3.4 In short, an "older" set and a "newer" set of projects would probably each show a higher M&E content in appraisal design than the sets examined here.

3.5 Third, the SAR does not always reflect accurately the M&E/KPI content of the project. In the new set, a few projects have appraisal reports which are generally "quiet" on M&E/KPI, but subsequent interviews revealed an incipient M&E/KPI plan had already been discussed and sometimes been described in background and other working papers. Also, some of the best of the new project M&E designs do not give any explicit KPI, but the interviewees revealed that they did plan to establish KPI, that these would be prepared later, and that the delay would allow beneficiaries to be more fully involved. "Best practices" in participatory evaluation call for beneficiary involvement as early as possible. In these cases the task managers saw an advantage in delay. Finally, and especially in the infrastructure sectors, monitoring and indicators are often part of the project routine: task managers expect them to be included in the quarterly progress reports or demand they be made available to Bank field supervision. Many of these task managers ignored or were unaware of the

17. "Issues in the Monitoring and Evaluation of Rural Development Projects: A Progress Report." Anderson, Dennis. RORSU. March 1976.

18. "Improving the Efficiency of Monitoring and Evaluation of Projects." OPS/AGRME. December 1985

Bank's directives to bring the indicators forward to the SAR.¹⁹ That behavior has adjusted radically to the post-PMTF era.

3.6 Fourth, in the following section the term "M&E" is used in its narrow sense, referring to programs of monitoring and evaluation activities established at the project level. The monitoring activities include indicators, but the emphasis of "M&E" is on the institutional framework to collect and use them and not on the KPI per se. The term "KPI", on the other hand, refers exclusively to the presentation at appraisal of indicator lists, whether or not the SAR also plans special institutional support to provide them. Viable KPI require that at least a rudimentary MIS system be operating - to supply the data. But they do not necessarily depend also on a capability for M&E, except where the indicators seek data that extend beyond the agency's routine flow of information.

B. Characteristics of the Old and New Projects

3.7 **Low Levels of M&E.** The first observation is the low frequency of significant M&E activity planned for the projects of either set. The 1989 OD called for effective M&E in all projects, but this mandate has been respected in less than half the projects where strong M&E should have been installed. This is shown in Table 3.1, which gives data on the content of M&E in the SARs for ten sectors. OED rated the M&E content at four levels - none, low, modest, substantial - where the last implies adequate treatment in SAR texts and annexes consistent with the requirements from monitoring and evaluation for fully tracking progress and analyzing impact. These are qualitative OED judgements on ratings, but the large number of SARs reviewed showed clear differences in the M&E content and allowed a differentiation at least between the "none/low" categories and the "substantial" category.

Table 3.1: M&E Content of Appraisal Reports: by Sectors

	Old Projects					*	New Projects				
	0	1	2	3	Total		0	1	2	3	Total
Agriculture	2	5	10	4	21	5	3	4	8	20	
Education	7	1	3	-	11	2	-	3	5	10	
PHN	3	2	3	3	11	-	2	2	6	10	
Urban	2	6	2	-	10	4	-	3	3	10	
WSS	6	-	-	-	6	-	1	2	2	5	
Industry	5	-	-	-	5	4	-	-	1	5	
Transport	10	-	-	-	10	1	4	2	3	10	
Oil & Gas	3	-	-	-	3	3	-	-	-	3	
Power	2	2	1	2	7	1	2	2	-	5	
Telecom	2	2	1	-	5	-	-	3	2	5	
Total	42	18	20	9	89	20	12	21	30	83	

* 0 = No mention (other than mid-term review, completion report and/or KPI).

1 = Low, pro-forma mention.

2 = Modest treatment, considered inadequate for effective M&E.

3 = Adequate or substantial M&E.

19. This behavior is partly explained by the pressure to reduce the size of the appraisal report. For many of these task managers, the effort to develop a short list of indicators and incorporate them in the SAR is an unnecessary reminder for the Borrower and a superfluous gesture for Bank management. The problem for management is that it does not know which task managers are hiding quiet KPI and which have none at all.

3.8 For example, in agriculture four of the old and eight of the new projects gave substantial importance to M&E. Thus there has been an increase in the percentage of high-profile M&E components (para 3.25). Nevertheless as percentages of the sector samples (19 percent and 40 percent) these figures are lower than would have been expected from the pervasive M&E culture usually associated with the sector, or from the exigencies of the operational directives. Conversely, seven of the old and eight of the new gave no, or only minimum attention to M&E. The pattern is repeated across the sectors. For the urban sector, the percentage of the samples with substantial M&E profiles increased from 0 percent to 30 percent. Recall that the agriculture and urban sectors were the two sectors provided with handbooks on M&E design (para 2.7). For all ten sectors, the figures are 10 percent and 36 percent.

3.9 **Sectoral and Regional Variations in M&E.** Sectors with higher M&E activity can be distinguished. Agriculture was the leader in the early 1980s, but that position is now disputed. Education, PHN and water supply/sanitation (WSS) have advanced relative to agriculture. On the other hand, the infrastructure sectors lag well behind these leaders, although there are some interesting exceptions at the project level.

3.10 On a regional breakdown, sharp distinctions in the old set have almost disappeared. This is shown in Table 3.2, which compares the regional totals of all sectors. The figures for the old Africa projects - 16 without any and one with high M&E content - can be compared with the even spread for South Asia - at three and three. Apart from South Asia, however, all of the Regions show falling ratings as one progresses to higher content levels. For the new set, the situation changes dramatically. With one exception, all Regions now show rising rates. Europe and Central Asia is the exception, with a concentration in the low ratings. No other significant differences emerge. High level M&E activity reaches across Regions as well as sectors: agriculture in India, Egypt and the Philippines and livestock in Ghana, education in Nigeria and Brazil, water supply in Indonesia, Sri Lanka and Brazil, railways in China and Hungary, telecommunications in Tanzania, etc. This is just a sample. Some of the high profile M&E designs are described in boxes elsewhere in the report.

Table 3.2 M&E Content of Appraisal Reports: By Regions

Regions	Old Projects					New Projects				
	* 0	1	2	3	Total	0	1	2	3	Total
Africa	16	7	6	1	30	7	4	5	11	27
East Asia & the Pacific	7	2	6	3	18	4	4	5	6	19
Europe & Central Asia	2	1	-	-	3	3	4	-	1	8
Latin America & the Caribbean	8	4	4	1	17	2	-	5	5	12
Middle East & North Africa	6	2	1	1	10	2	-	4	3	9
South Asia	3	2	3	3	11	2	-	2	4	8
Total	42	18	20	9	89	20	12	21	30	83

* 0 = No mention (other than mid-term review, completion report and/or KPI).

1 = Low, pro-forma mention.

2 = Modest treatment, considered inadequate for effective M&E.

3 = Adequate or substantial M&E.

3.11 Even within a country program there is no consistent pattern. Brazil seems to exhibit a relatively stronger propensity for serious M&E: that is shown in a new education project (Box 5), a two-year old water supply project (Box 16), and a completed rural electrification project. But the Brazilian portfolio has another proposed water supply project with relatively weak M&E, and a new urban project with an equally undistinguished M&E component. (A difference between M&E performance in the northeast and northwest rural development programs is presented in Boxes 7 and 10 and discussed in the next chapter). The relative strength of Brazil in the M&E sample is partly explained by the acknowledged intellectual vigor of staff in at least some of the federal and state ministries and parastatals, where Bank innovators found willing collaborators in the search for evaluative tools. The federal structure of government may itself have something to do with the planning vigor at the state level. The Philippines is another special case. There one finds respectable M&E in many sectors, and a national commitment to M&E in the planning ministry. Philippine consultants are exporting M&E to other countries.

3.12 **Use of Indicators.** Staff incorporated comprehensive KPI, including at least a few impact indicators, into 1993 appraisal design at the same rates they did M&E systems (36 percent for the whole sample). The rate was well below the level reflecting satisfactory compliance with the OD (the KPI requirement should have been easier to respect than that for M&E). The level of KPI activity in the older set was much lower: again about the same as for M&E (12 percent). This is shown in Table 3.3.

Table 3.3: KPI Content of Appraisal Reports

	Old Projects					New Projects				
	* 0	1	2	3	Total	0	1	2	3	Total
Agriculture	18	2	-	1	21	8	3	2	7	20
Education	7	1	2	1	11	1	2	3	4	10
PHN	7	1	-	3	11	2	2	3	3	10
Urban	3	5	1	1	10	3	-	2	5	10
WSS	1	4	1	-	6	-	-	2	3	5
Industry	5	-	-	-	5	3	1	1	-	5
Transport	5	1	1	3	10	1	3	1	5	10
Oil & Gas	2	1	-	-	3	3	-	-	-	3
Power **	2	2	1	2	7	2	1	1	1	5
Telecom	-	-	5	-	5	-	-	3	2	5
Total	50	17	11	11	89	23	12	18	30	83

* 0 = No indicators (They are implicit in other tables).

1 = A few input and process indicators.

2 = Extensive input and process indicators and some output indicators.

3 = Extensive KPI, including output and impact indicators.

** Exclude Environmental Impact Assessments.

3.13 In agriculture, the growth trend of high-content cases reported above for M&E is repeated for KPI. Nevertheless, the low numbers are truly surprising. Only 1 of the 21 old SARs had either an indicative or an illustrative KPI table in its text or annexes. That statement refers narrowly to

tables with indicators identified as such. Of course, most of these projects had project implementation tables, and most also had rates of return analyses where the indicators of implementation progress were self-evident. So the reference here is to the practice of extracting and identifying as such a subset of "key" indicators which can be used summarily by project management and by Bank supervisors to assess overall progress. Thus in the early 1980s the authors of this sample of "old" agriculture SARs did not consider it essential to identify the key indicators in those plans, even though supervision was supposedly still obliged periodically to fill them in its reports. For the new agriculture set of 20 projects, only 7 (40 percent) new SARs provide indicators - some a single table of a few important input and output variables, some with many pages of indicators in an annex.

3.14 For education, PHN, urban and WSS, the KPI are better - 13 percent and 43 percent for the four sectors taken together. Under this heading the urban performance in KPI repeats the expansion in M&E, growing from 10 percent to 50 percent.

3.15 The other five sectors in the table are the "infrastructure" (or "hard") sectors, whose ratings can be compared to those in the five "social" ("soft") sectors just reviewed (the "social programs": note that urban and WSS are included in this group for purpose of this analysis). Measured for M&E content, the infrastructure sectors run significantly lower than the rest; measured for KPI content, the difference is smaller.

3.16 But the use of KPI to track physical performance in the utilities and other infrastructure sectors varies according to the nature of the product targeted by the project. For some sub-sectors, there has been substantial progress in the sophistication of the tools of measurement. The models and methods for assessing the state of highways and pavements, in order to measure utilization levels and rates of deterioration of roads, as well as to determine the requirements for maintenance and rehabilitation, is the most sophisticated exercise involving indicators in the Bank (Box 1). In telecommunications and ports, the physical KPI were standardized decades ago. In telecom, indicative KPI have been available in that group's technical library since the late 1970s (in the Telecommunications Technical Notes series). Telecom was one of the first sectors to deliver sector-wide indicators to the ongoing OPR task force. Since in this sector the character of the portfolio is shifting, reflecting the increased attention to privatization and cost-based pricing, the list of indicators has been expanding to cover the new criteria. For ports, given the high costs of maintaining a ship in the harbor or at the dock, unloading and turnaround indicators have been standard practice since before the Bank entered the field. Electric power projects have their own set of physical indicators: "unaccounted-for-losses" being one of those traditionally emphasized by this profession.

3.17 With few exceptions, the infrastructure projects do not call for KPI that go beyond measurement of project inputs and a small number of direct outputs (e.g. ship turnaround time and unaccounted-for-losses of power). The number of these output indicators on the KPI list has expanded since 1983. But higher order outputs and final or indirect impacts have rarely been targeted. Thus, even when the SAR describes expected benefits due to expanded economic activity attributable to the project, and puts consumer or producer surplus into the economic rate of return (ERR) analysis, project supervision staff are not expected to update these economic projections against monitorable KPI targets.

3.18 This is the case, for example, with two new Chinese railway projects, one of which (Sixth China Railway) OED considers a good example of a substantial monitoring plan (Box 4). Both SARs use the consumer surplus attributable to declining transport prices to derive the benefits; one (Inner Mongolia Railway, outside the sample) uses generated traffic as well. But neither specify KPI on cost savings to railway users or incremental economic activity. The task manager felt it would have been

worthwhile to monitor the increase in economic activity, but that would have required a stronger borrower commitment and closer Bank-borrower dialogue.

Box 1: Indicators for Managing Pavements

The following passages are taken from a paper prepared by Bank technical staff for the Third International Conference on Managing Pavements, scheduled for May 1994.^{a/} They reflect the concern of the Portfolio Management Task Force, and an intention to put indicators to work:

"This framework for defining performance indicators for the road sector has been developed in the broader context of the modern trend to monitor performance and increase the accountability of those providing services and responsible for public expenditures. In the specific context of the task of managing pavements, the framework is particularly apposite. Pavements are one component of the infrastructure provided for transport, and the indicators identified here are of relevance and interest to the various groups involved, namely the private and commercial road users; the policy-makers and regulators; and the many groups of providers (including the highway agency for planning, programming, budgeting and implementation services; the consulting service agencies for design, supervision, etc.; the suppliers of goods and materials; and the funding or financing agencies."(page 17).

The paper identifies a number of specific performance indicators to satisfy those goals, and defines them sufficiently to indicate what information is needed. It stops short of specifying all of them quantitatively and unambiguously, "partly because this is still in progress but also because of the need for consultation in the sector to reach consensus". The categorization of the indicators distinguishes "infrastructure" (road lengths, types of surface, vehicle fleet, etc.); "service quality" (road surface ride, noise ratings and skid resistance, curvature and gradient, average speeds, etc.); and "process" indicators (measuring the efficiency and effectiveness of construction work, job backlogs, revenues and expenditures). The variable "pavement roughness incidence" is one of the more important on the list, because under modern road planning practices it determines the sequencing of repairs.

The paper emphasizes reliability and cost factors that help determine which indicators should be collected under given conditions:

"To maintain the consistency of performance data for multiple uses and users, the reliability of the data is important. Reliability is determined from the accuracy (in terms of precision and bias) of individual data items, the sampling frequency in space and time, the type of data aggregation techniques used, and the currentness and temporal coverage of data items used in projections. Choices among the various factors mentioned above will depend on the final use to which the data will be put and the resources (financial and staff) available to collect, maintain consistency, and keep the performance data current" (page 9).

^{a/} "A Framework of Performance Indicators for Managing Highways and Pavements." Humplick, Franjie, and Paterson, William D.O. Transportation Research Board, #PM002. 1994. This paper follows a series of reports produced in the Bank in the last decade aiming to systematize the analysis of national road networks, their costs and benefits, maintenance standards, and indicators determining priorities for expenditures. One other important recent report in this series is: "Information Systems for Road Management: Draft Guidelines on System Design and Data Issues." Paterson, William D.O., and Scullion, Thomas. Infrastructure and Urban Development Department Technical Paper. September 1990. Report #INU 77.

3.19 The same is true of rural roads projects. Those that rehabilitate or seal roads already bearing high density traffic base their ERR calculations on savings in vehicle operating costs (VHO). Those that improve lightly used roads, or extend roads into new areas, base their ERR calculations on generated economic activity. Neither group specify KPI to track either VHO or generated economic activity. They stop at routine measures of direct output - kilometers resurfaced, etc.

3.20 Rural electric power (RE) is a subsector with special indicators. Since most of the RE group have been financial drains on the public utility, other methods for measuring project impact have been employed, including surveys of household uses of the energy. This is one of the few examples where the utility projects have added an M&E capability precisely in order to identify welfare effects that eluded the financial tests (see footnote 40 to para 6.16).

3.21 One exception to all the other comments on KPI is the oil and gas sector. In the six SAR's in the sample - three old and three new - none made any reference to either KPI or M&E. This is a prime example of the quiet KPI phenomenon. But the absence of *any* appraisal reference to KPI suggests that indifference to the directives can be carried too far.

3.22 **Role of Financial Indicators.** The most frequently reported KPI are indicators demonstrating the state of financial health of the project operation and of the individual implementing agencies. These are included in the KPI data in Table 3.3, and often dominate the SAR KPI lists. For "credit" projects - in agriculture, industry, housing, etc. - financial indicators are usually the only KPI called for by the SAR.²⁰ Financial KPI have been standard practice in old as well as new projects across all the industry projects and for telecommunications, power and other utilities.

3.23 Furthermore, for the industry and utility sectors, financial KPI are considered by Bank professionals to be the essential test of project success. Profitability of the enterprise is interpreted as the paramount measure of product acceptability. Thus, Bank staff are not expected to go on to assess consumer uses of and satisfactions with individual products and services. Since financial indicators are derived directly from routine MIS financial data, the other institutional requirements of M&E disappear.

3.24 This position has been challenged on at least two counts. First, it is weakened by the common practice of public subsidization of consumer utility tariffs. Subsidies imply distortions in consumer uses and satisfactions. To that extent the assumption that the market is a substitute for an evaluation program no longer holds. Second, the position is weakened wherever there are important externalities. Nevertheless, the closer the administered tariff approximates market values, the more legitimate the claim that projects in this sector need not develop an internal M&E capability. The broader effects must be examined, but that job is better handled by special evaluation studies rather than internal project M&E procedures. The point here is that some projects with good financial indicators but without conventional M&E may nevertheless satisfy the primary objectives of project M&E.

3.25 **Recent Burst in Good M&E/KPI Design.** Tables 3.1 and 3.2 show that the rate of compliance with the mandatory instructions of the two ODs has been inadequate. Nevertheless, the tables show as well there has clearly been an increase in M&E content in the SARs in the approximately ten-year period between approval of the two sets of projects. The percentage of SAR's with substantial M&E moves up from 13 percent to 36 percent, and those with substantial KPI move from 12 percent also to 36 percent. Interviews with task managers of new projects revealed that some other projects rated low on the basis of the SARs in fact had an implicit M&E orientation that was detectable in the scheduling of project inputs although it had not been described as M&E. Two interesting examples from Indonesia are discussed in para 3.32. There were no comparable

20. That was not the case in agriculture in the early 1980s, but the Bank subsequently abandoned reporting on physical progress of credit projects at the farm level.

examples in the set of old projects. The conclusion is that M&E concepts are indeed spreading, though not yet as widely as would be desirable.

3.26 But the more important observation from the new set of SARs, including the extra projects with high-profile M&E, is that the ones with the highest M&E and KPI content are the haphazard product of spontaneous action by isolated individuals with a personal predisposition to put the tools of monitoring and evaluation to work. Division chiefs appear to have little to do with the *emergence* of this disparate group, though they *encourage* the innovators to continue. Small fires are burning in all of the social program sectors, and in some of the infrastructure sectors as well. Examples of SARs with high M&E content drawn from both the soft and hard categories are described in boxes – India Uttar Pradesh Sodic Lands Reclamation (Box 2), Yemen Education Sector Investment (Box 3) and Sixth China Railway (Box 4). These task managers are unorganized as a promotional force for M&E, and either unaware of or indifferent to the ongoing search for sector-wide KPI. They are simply "doing their thing", the initiative is theirs, the division chiefs are supporting them, but if together they had reduced the M&E or KPI content the task managers in most cases believe that the adjustment would have gone unnoticed.

Box 2: INDIA - Uttar Pradesh Sodic Lands Reclamation Project
Cr.2510-IN for US\$54.7m - Approved May 1993

This is a new project making extensive use of M&E. On the Bank's part, it is the work of a single task manager - old in the Bank, new to the India division, and without experience in designing an M&E component. The intention to ground the project design to M&E was his own. There was little prompting from the Region. His previous assignment was on Nigeria, and included supervision of the ADPs, including their problematic M&E operations (para 4.25) which he was determined not to replicate in subsequent appraisal work. The project will reclaim 45,000 ha that have been forced out of farming by salt accumulation. Shallow tubewells will be used to flush the salts and irrigate about 4 ha each - roughly six farmers per well. Organizing farmers around communal wells is a new experience in this state, which will use the project as a pilot for additional reclamation operations. The task manager started from the principle that the farmers had to be closely involved in project implementation, and participatory features are strong in appraisal design.

Four M&E programs will operate side by side under at least three agencies. The state Land Development Corporation (LDC), the principal executing agency, is responsible for monitoring input and process indicators. The Remote Sensing Application Center (RSAC) is responsible for monitoring physical changes of land and groundwater conditions, based on both satellite imagery and ground work. This includes an assessment of project-induced effects on regions bordering the project. The Indian Institute of Management, Lucknow (IIM), has two jobs. It will monitor the physical outputs of the project - wells, farmers, cultivated areas, etc. - taking over that responsibility from LDC (IIM has already completed for LDC a proposal for a project MIS system and annual planning program). It will also be responsible for socio-economic surveys and analysis of the impact of the reclamation program and the 4 ha well technology on production, farms and incomes. The task manager has structured the implementation program to advance by annual steps, to take maximum advantage of lessons from the M&E activities. The Joint Managing Director of LDC reportedly is committed to the M&E format. The task manager rates the project high for government ownership.

3.27 The 1989 ODs which mandated M&E and KPI may have influenced the shift to higher levels of activity, but this is not obvious from the interviews. A small minority of these task managers knew there were directives on both M&E and KPI. The project advisers in the departmental front offices played a definitive role in several of the cases discussed in the interviews: for example, they accompanied the appraisal teams to Mexico and Yemen in support of the new Mexican Highway Rehabilitation and Traffic Safety Project and Yemen Education Sector Investment Project - both with good M&E design, and involved themselves with securing government ownership of the M&E

component. The Project Advisor position is one of the important routes for transmitting the messages from the ODs (and PMTF). But it is not yet a decisive force, at least up to the last half of 1993.

3.28 There is one exception to the haphazard character of the spread of M&E. The Water & Sanitation Division of TWU has created a new model for low-income targeted communities which has been put into practice in recent projects managed by the Bank's Brazil and Indonesia infrastructure staffs. TWU is seeking now to promote the model in other countries and other sectors, under the term of "structured learning" (SL). M&E is essential to the process, and SL will carry M&E with it wherever it goes. SL is discussed further in Chapter V.

Box 3: YEMEN - Education Sector Investment Project
Cr.2570-YEM for US\$33 Om - Approved February 1994

This SAR, supported by three implementation volumes, presents one of the most extensive sets of KPI found in the OED survey. The objective of the project is to assist the government to implement policy reforms in the education sector, addressing in particular the secondary and post-secondary technical levels. The objective is to be achieved through a "Planned Change Program", working within eight so-called "change areas". These include physical facilities, teacher performance, curriculum development and female access. Each "area" is divided into subprojects, and for each of these subprojects separate lists of input and output indicators are given. The indicators refer to quantitative and qualitative variables; wherever possible the former are given numerical targets. The SAR provides the lists and single line summaries of the KPI. Implementation Volume 1 provides additional detail. The SAR also lists in a separate annex, without targets, impact indicators covering numerous "efficiency" and "quality" factors. The annex requires base line, concurrent and post-project surveys. Institutional responsibilities are assigned for each of the indicators. A mid-term evaluation is required, with exhaustive examination of trends to date.

What sets this SAR apart is the intensive participatory process over two years that led to the proposal. Working groups of governmental officials, headmasters, teachers and professionals were formed for each "change area". The task manager, a veteran of six years of involvement with Yemeni education, not only pushed the groups to develop their own objectives and strategies for change, but to articulate what intermediate and final results were expected and to express them as indicators. He held seven workshops during the preparation period. He employed international and local consultants, but his instructions to them were to guide the process and neither impose their ideas and objectives nor suggest the indicators. He is satisfied that those instructions were reasonably well respected. If so, this SAR embodies one of the most participatory approaches to M&E design identified in the study.

3.29 Whether the more elaborate M&E operations planned in the new set will succeed is uncertain. Few of the SARs give any sense of government "ownership" of these components. Subsequent interviews revealed a high level of participation by government and beneficiaries in most - but not all - of the projects, and these assertions by the staff about country ownership are convincing. For example, seven of ten interviewees on new urban projects claimed high levels of government ownership of M&E, and nine of eleven of the PHN interviewees said so too. Nevertheless, the agreements are untested in practice, and *that* story, told in the next chapter, shows very low levels of ownership during the 1980s.

3.30 There is also a capacity-building problem - bigger in some sectors and regions than others. Expectations of M&E are not supported by adequate inputs into M&E from the project and the Bank. The M&E products, including KPI, are expected to materialize partly on faith and good

intention. For reasons that were not clarified in the interviews, the urban sector seems especially prone to this problem: a mismatch between M&E plans and resources. Simply backing up the M&E targets by prior agreements with government - as announced in the SAR section on "agreements reached" - has not guaranteed results in the past.

Box 4: CHINA - Sixth Railway Project
Ln.3581-CHA for US\$420.0m - Approved March 1993

This large sector-modernization loan finances electrification of two main lines, installation of a modern, system-wide telecommunications network, a mechanized system for preventative track maintenance, modern machine tools for manufacturing critical locomotive and rolling stock parts, a pilot container transport program, and technical assistance. The project makes exceptional use of monitoring indicators, compared with the norm for the transport sector. However, it resembles the rest of that group in not calling for evaluation studies to test the assumptions about economic impact. There is a minor evaluation component related to resettlement of small communities displaced by the power substations to be built to support line electrification.

What is most impressive is the attention given to identifying appropriate indicators. For each of the six project components, the Bank requested specialized groups within the Ministry of Railways to develop their own lists. The initiative was taken by the Bank, but by turning the responsibility over to MOR project staff the monitoring program has been effectively internalized. The Bank's consultants divided the indicators into two sets, called performance indicators and monitoring indicators. Each of the first set carries specified annual targets, which government has agreed to respect. The second set - much larger than the first - lists other important checks but without numerical tests. As an example of the first set, the single performance indicator for the track maintenance component is the number of locations on each stretch of 100 km, of those rail lines that are converted to mechanized maintenance, that require emergency repair in a year. The target starts at 59 in 1993, and declines progressively to 25 emergencies per 100 km in 1996. The purpose of that performance target is to serve as a proxy for an overriding objective - to keep all lines which use mechanized maintenance in good repair. Another example is from the several performance indicators assigned to the locomotive and rolling stock component. In this case the test of project effectiveness is defined by indicators of failures, availability, and reliability of locomotives and wagons - comparing the older units with those manufactured using the high-tech tooling equipment. In indicator terminology, these are all examples of higher-order outputs. They do not measure the impact of the project on economic activity. The Bank did not think that was an appropriate component of a project M&E system, primarily because economic activity was considered to be outside the control of the implementing agency.

3.31 **Some Other Characteristics.** There are only four examples among the 83 new projects where the SARs anticipate the PMTF recommendation calling for annual KPI *threshold targets*, to be used to identify poorly-performing project components subject to review, redesign or cancellation. But again, usually there is no information on government "ownership" of this aggressive application of KPI. Three of these exceptions appear in country programs on the African continent - Algeria urban, Zimbabwe PHN, and Tanzania telecommunication. Government's full acceptance of the tough conditionality implied by the indicators has yet to be demonstrated. The other exceptional use of KPI targeting is in the two new Brazilian Basic Education Projects for the Northeast states. Annual releases of Bank funds for school construction are conditioned upon each state government reaching annual performance targets for three policy variables. OED examined this case in the field and is satisfied that the present governors of the states are indeed "on board". (There is an air of confidence also that the new governors to be elected in late 1994 will have to accept the commitments.) This is another one of the high profile M&E (and KPI) examples, and is described in Box 5. One recommendation of this report is that the process of intense and prolonged discussion that led to the agreements on thresholds indicators in Brazil be repeated in other projects.

Box 5: BRAZIL - Second and Third Northeast Basic Education Projects
Ln.3604-BR for US\$212.0m - Approved May 1993; Ln.3663 for US\$206.7m - Approved November 1993

These projects exemplify "best practice" for deriving targeted indicators to achieve policy reform within the context of an investment project. The Bank, and most professional educators in the state ministries, wanted to focus the proposed project on improving the quality rather than expanding the infrastructure of educational services. But construction was politically attractive. To ensure state government commitment to the quality dimension, the Bank agreed to finance physical rehabilitation and upgrading of classrooms and other facilities. But it insisted that a short list of key quality improvement measures be embodied in verifiable annual targets to be met by each state. To that extent the targets were imposed. But the Bank worked with the nine northeast states (and the federal ministry) over a period of three years starting in 1991 to identify the fundamental problems explaining the inefficiency of the educational system, and to build a consensus around a short list of meaningful indicators. Actually the project was first prepared in 1985, at a time when Brazil and the Bank were hoping to split the POLONORDESTE multi-sector rural development program into separate, sectoral lending lines. Agriculture was broken out as PAPP in 1986 (see Box 10), but successor programs for the health and education sectors were delayed. In 1991 the Bank received Japanese technical assistance funds to support a major effort to design a quality-oriented program for the first four years of basic education. A decision was made to divide the new education proposal into two projects, the first to pick up four of the nine northeast states and the second to pick up the other five.

In 1992 a team of Bank and ministry staff and their consultants took the lead in dimensioning indicators that would determine whether the Bank would disburse against expenditures on infrastructure - for that state in that year. One indicator is the ratio of students to employees in the state-run primary school systems. The other is the percentage of the recurrent education budget spent on didactic materials for students. This dimensioning process was as important as selecting the indicators. Agreement was needed on the exact definition of terms so that the measures were unambiguous: how to handle seconded and part-time teachers, for example. Agreement was also needed on the numerical targets: the end-of-project targets and the shape of the curve of annual targets to get there from the year 0 position (i.e. convex, concave, or straight line). The deliberations were not hurried, and the states' self-interest in getting sensible and constructive targets was recognized. OED met with state government officials in three of the states. It is clear that the governments have indeed internalized these "aggressive" indicators. State education officials can quote the annual targets (one carried a pocket calculator with a program displaying the targets as well as measures of progress to date). State officials use the same targets to reject petitions from groups seeking new appointments to the staff. It is clear that the long gestation period was essential to building this coincidence of interests around targeted indicators which in other circumstances could have been treated as Bank-imposed and expendable.

3.32 A similar strategy for putting indicators to work to force the pace of project implementation was observed in two projects in Indonesia, one for rehabilitation and improvement of rural provincial roads and another for completing irrigation works left unfinished by earlier projects. These are unusual examples of M&E features, because the SARs do not call them by that name and they were rated poorly for M&E during the desk review. In these cases, the *disbursement schedule is deliberately divided* in two phases. The first, of about a year, is treated as a testing period to ensure that commitments are being respected and that implementation follows design. This is a tranching mechanism, but linked, not to compliance with policy objectives (as in the Brazil education projects discussed in the last paragraph, and in adjustment operations), but to lessons from start-up activities. These two initiatives also have yet to be proven effective. The irrigation project is disbursing extremely slowly. But the concept of building a demonstration phase into the disbursement schedule can be considered a type of structured learning operation. The two projects are not in any sense early and deliberate pilots of a systematic shift by the Indonesia department to a new "learning" model for the project cycle. That may yet happen. Rather, they are the spontaneous expression of two individual task managers who wanted to ensure their projects improved with experience. They did not know what other task managers were doing. Nevertheless, their example may show the way to

a new, phased modality for the project cycle, at least for projects with uncertain prospects at appraisal.

3.33 There are only a few examples also of *participatory evaluation* in M&E, where that implies monitoring by the beneficiaries themselves and their participation in project redesign, and sometimes implies as well beneficiary participation even in the initial design of the M&E system. OED counted 4 in the new agriculture set and 3 in other sectors. One remarkable example is described in Box 6, the Egypt Matruh Resource Management Project, which intends to involve the coastal Bedouin in monitoring their own progress.²¹ PE appears in some other new projects outside the sample. It is too early to determine whether this new concept, popularized in recent Bank publications, will extend broadly. It is inherent in the structured learning approach, and would seem to be an inevitable step for projects moving into the softer social areas (para 6.16).

Box 6: EGYPT - Matruh Resource Management Project
Cr.2506 for US\$22.0m - Approved May 1993

This project deals with the Bedouin tribes settled on the low-rainfall margin of the northwestern coast. The purpose is to help them conserve the meager natural resource base of the coastal zones while exploiting the potential for rainfall water harvesting and other agricultural activities. Tribal structure dominates social and political organization, and the project would work closely with tribal representatives to design, implement and monitor the program. There are about 40 recognized tribes - subdivided into clans - in the area. For each a Community Group (CG) is being organized, responsible for elaborating and executing a three-year Community Action Plan (CAP).

Local participation has been the distinguishing characteristic of the preparation period, and will dominate implementation. CG representatives will work with the extension service and other public agencies involved with project execution. Annual progress reviews are planned, bringing representatives of all these groups and agencies together to discuss problems, solutions and revised plans for the next year. But the more unusual part of the design is to allow the CGs to monitor their own progress - in establishing erosion controls, expanding the orchards, harvesting rain-water, taking better care of their herds of sheep and goats, etc. The Memorandum of the President (there is no SAR) has an 11 page annex devoted to "monitoring and evaluation indicators". It is divided in five sections: input indicators, output indicators, beneficiary contact monitoring, evaluation of effects, and impact evaluation. Participatory Rural Assessments were carried out during preparation, and a Participatory Rural Needs Assessment will also be conducted to develop a more detailed baseline for subsequent evaluations. NGOs and/or universities are expected to be involved in the beneficiary contact monitoring, which in this M&E structure refers to visits by professionals to the households. CG involvement in self-monitoring refers more to the input and output indicators. The Project Management Unit will have an M&E specialist, and he will be supported by four M&E officers - one at each of the four Sub-regional Support Centers. All this M&E work at ground level will be complemented by satellite imagery, to get information on improvement in the watersheds and expansion of agriculture. In OED's survey sample of new projects this ranks among those with highest M&E content, as well as distinguishing itself for allowing a substantial element of participatory evaluation. "Allowing" does not do justice to this design: the task manager is one of those Bank staff who feels success depends on it.

3.34 A few of the more important observations on *MIS design* are worth reporting. It was easy to identify projects with a large Bank contribution to MIS - to introducing, expanding and computerizing information systems. But there were many projects where an adequate MIS was

21. This project and its participatory features are described in the leading article of a recent issue of Bank's World "The Matruh Resource Management Project, Gaining by Losing Control." Souhial, Bachir. Bank's World, Volume 13/Number 5, May 1994.

already running, where previous or associated Bank projects had already financed development of MIS, or where other donor agencies were already providing grant assistance. In these cases the Bank considered further support unwarranted, and the subject was ignored in appraisal documents. Nevertheless, it is clear that the Bank significantly expanded its support for MIS during the decade 1983 to 1993. The shift was common across all sectors, although it was most prominent in agriculture and education. This reflects the low base of MIS activity in these sectors at the beginning of the period, and they still lag far behind the infrastructure and industry sectors in the quality of the MIS. In infrastructure, recent developments in telecommunications and the transport sector (for both roads and railroads) have propelled the Bank to the cutting edge of MIS computerized systems. These are being introduced to a growing number of borrowers. The pavement indicator report mentioned in para 3.16 and Box 1 is linked to the Bank-developed, universal MIS road model.²² In the power sector, adequate MIS is now installed in most electrical energy production and distribution agencies in borrowing countries. There is less room for the Bank to promote further development.

3.35 Finally, in almost all projects where the Bank provided support to MIS, the SAR linked MIS only by implication to whatever KPI or M&E components were also proposed. MIS was promoted in order to enhance management and financial control functions. This is an interesting distinction from KPI, since it shows the Bank at work with MIS to strengthen project and agency management, and not aim it at satisfying Bank information requirements as is the case with some of the KPI activity.

22. The Highways Development and Management Model (HDM), the first generation of the highway management modelling schemes that are now in use.

IV. PROJECT M&E IN PRACTICE: DISAPPOINTING RESULTS

A. General Remarks

4.1 After studying appraisal designs, OED looked for results: the performance of M&E programs in the old set of projects. Unfortunately, the institutional memory problem was severe in interviews on the old set. These projects were completed about two to three years ago, except for those whose selection was triggered by the issue of a PAR instead of a PCR (which makes them older). At least half of the PCRs were prepared by consultants, overseas mission staff, or headquarters staff who either came into the project just before completion or subsequently switched divisions. Since the great majority of these PCRs give little or no information on M&E anyway, and those that present KPI in tabular form do not explain how the figures were derived or the sources, they provide little insight into the extent of actual monitoring and evaluation activity. Asking staff now to remember what was going on in the M&E arena that they had not bothered to report on before was usually futile. Thus one of the interesting hypothesis at the beginning of the study - that the well-managed and successful projects were likely to be supported by effective monitoring, whether it was reported in the Bank's ex-post evaluation documents or not - could not be tested. Presumably supervision and progress reports would have shed more light on the subject, but these were excluded from the desk review (para 1.11).

4.2 Table 4.1 presents OED's rough measures of the quality of the project M&E work, as discerned from the PCRs and PARs. Of the four columns, only the right hand one represents M&E activity at effective levels and equal to or better than those proposed at appraisal. The other columns represent - from right to left - M&E of some value but less than anticipated at appraisal, minor M&E and no M&E (or no reporting of it). The data imply that in the five social sectors only 19 percent of the 59 projects had M&E rated as modest or better, and in the five infrastructure sectors the figure is 7 percent. Remember that industry and the other sectors grouped here under the infrastructure label cannot be assessed properly on the basis of this poor showing on M&E. Their ratings are better for the quality of KPI (not shown here). As mentioned above, *that* may accomplish the goal of good M&E (para 3.23).

Table 4.1: Quality of M&E in Practice

	Old Projects				Total
	* 0	1	2	3	
Agriculture	-	9	6	6	21
Education	7	1	3	-	11
PHN	-	7	1	3	11
Urban	2	5	1	2	10
WSS	6	-	-	-	6
Industry	5	-	-	-	5
Transport	9	1	-	-	10
Oil & Gas	3	-	-	-	3
Power	2	2	1	2	7
Telecom	5	-	-	-	5
Total	39	25	12	13	89

* 0 = No effective M&E or no mention.

1 = Low M&E activity.

2 = Modest M&E activity, less than planned and incomplete.

3 = Effective M&E.

4.3 Those M&E performance ratings can be compared with the ratings of the M&E content at appraisal. As shown in Table 3.1, 12 percent of the social projects in the old set were rated with "modest" or "substantial" M&E appraisal content, and 7 percent of the infrastructure projects. These numbers compare with 19 percent and 7 percent performance ratings, given in the last paragraph (but see para 4.16). OED then looked at the individual projects, to find the frequency of projects rated poorly or well in both respects. Figure 4.1 shows the results. A relationship between appraisal content and performance stands out. It is revealed by the strong diagonal of larger blocks descending from the 0/0 to the 3/3 coordinates. Despite the generally low numbers of projects with either substantial design content or favorable result, the graph demonstrates a clear, positive relationship, implying better appraisal work pays off. As discussed below, however, the case studies of actual performance of M&E in agriculture fail to confirm this intuitively attractive result. Massive appraisal inputs into M&E did not lead to acceptable results. One must use this chart with caution. Better appraisal makes for better results, but not necessarily anywhere close to the targets.

4.4 OED also looked for a relationship between the ranking of projects by M&E performance and the overall ratings of projects during supervision and at completion. If M&E is serving management as it should, and feedback and corrective mechanisms are in action, better M&E should show up in better projects. Whether the relationship is strong enough to upgrade project completion ratings from "unsatisfactory" to "satisfactory", or supervision ratings from "3" to "2", is unlikely, since good M&E is not expected to rescue bad projects. Nevertheless, the relationship should at least be positive.

4.5 For the completion ratings, that was the case: a weak but discernable positive relationship between good M&E performance and good completion ratings. This is shown in Table 4.2. Eighty-five percent of the 25 old projects rated as 3 in Table 4.1 ("effective M&E") were rated satisfactory at completion, while only fifty-nine percent of the 63 old projects rated as 0 or 1 in the table were rated satisfactory. Regional differences are not pronounced: the relationship is strongest in East Asia and weakest in Africa but all Regions show the same trend. Sectorally, the social sectors show a stronger relationship than the infrastructure sectors, though the latter set has only a few good M&E examples and the numbers are too small to have much meaning.

4.6 For the supervision ratings there is practically no difference between good and bad M&E performers. Table 4.2 shows that the average of all supervision ratings over a project's lifetime was not affected by the quality of M&E (in the OED sample of old projects). One could hypothesize that better M&E would improve project management, project implementation, and thus the supervision grade - the same effect that shows up in the completion rating. But there is an offsetting factor that may explain why the supervision ratings are unchanged. This is related to the so-called "disconnect" between supervision and completion ratings, the phenomenon that attracted Bank management attention in the early 1990s and led eventually to appointment of PMTF (para 2.18). This "disconnect" refers to the difference between supervision and completion grading. Supervision ratings were usually better than completion ratings, suggesting an element of over-optimism in the supervisor's assessment of progress. One could hypothesize also that better M&E would provide the supervisor with a more realistic basis for his ratings, tending to depress them. Thus, for supervision ratings, better M&E should simultaneously improve the project and improve the assessment, opposing forces (in the "disconnect" scenario) which could cancel one another out. Table 4.2 includes a measure of the disconnect, and it does indeed fall as M&E performance improves (from 23% for the poor M&E projects to 12% for the good ones).²³ But this search for a relationship between M&E and supervision is too speculative to carry much weight. The rapid "overview" exercise is not the proper analytical instrument to try to investigate the impact of M&E on rating behavior at either completion or supervision.

23. This is inevitable the way the "disconnect" is presented in Table 4.2: as M&E improves, completion ratings improve but supervision ratings remain constant. Note that the supervision rating used here is a lifetime average, not the final supervision report rating.

Figure 4.1: Quality of M&E in Old Projects:
Ex-ante vs. Ex-post

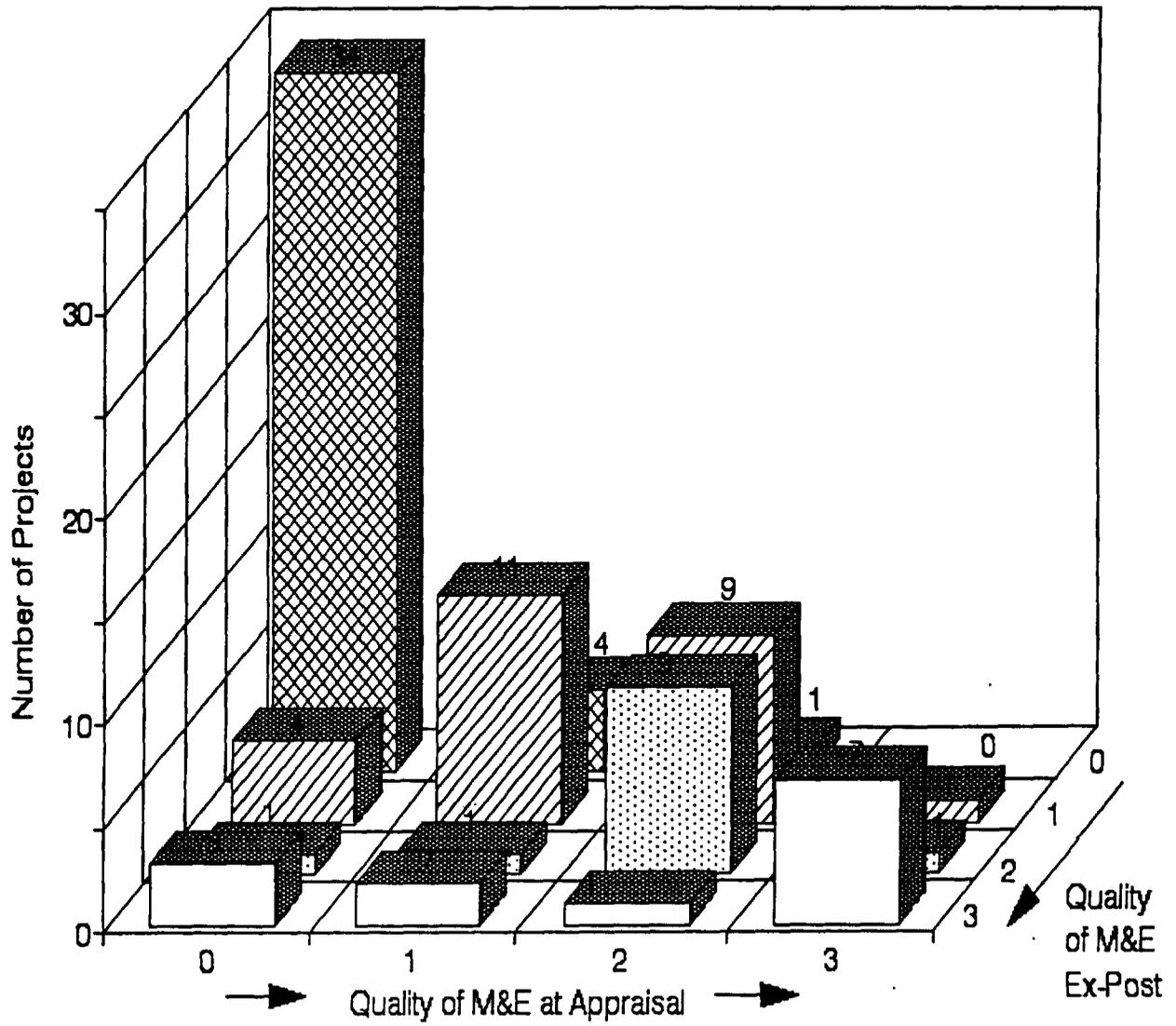


Table 4.2: Project Ratings during Supervision and at Completion
vs. OED's Ratings of M&E
(For 89 Old Projects)

	OED's Ratings of the Quality of M&E Components Ex-post		
	Low (0+1) *	Medium (2) *	High (3) *
Ratings during Supervision**			
Good	18 (28%)	3 (25%)	4 (31%)
Fair	41 (64%)	8 (67%)	8 (62%)
Weak	5 (8%)	1 (8%)	1 (8%)
Total	64 (100%)	12 (100%)	13 (100%)
Ratings at Completion			
Satisfactory	44 (69%)	9 (75%)	11 (85%)
Unsatisfactory	20 (31%)	3 (25%)	2 (15%)
Total	64 (100%)	12 (100%)	13 (100%)
"Disconnect"			
Percent Rated Good or Fair During Implementation	92%	92%	92%
Percent Rated Satisfactory At Completion	69%	75%	85%
Difference	23%	17%	7%

* M&E Ratings: 0 = No effective M&E or no mention; 1 = Low M&E activity; 2 = Modest M&E activity, less than planned and incomplete; 3 = Effective M&E. M&E ratings are drawn from Table 4.1.

** Supervision Ratings were considered "Good" for projects with an average rating of less than 1.5 during implementation; "Fair" implies an average rating of 1.50 - 2.49; "Weak" implies an average rating of 2.5 or more. For this table, an average was taken of all Form 590 Supervision ratings during the life of each project.

4.7 Data is not available to give an estimate for the aggregate cost of M&E in the portfolio. At appraisal, M&E was usually not itemized in the SAR cost tables (presumably most parts of the M&E bill can be derived from the appraisal working papers). That is less true with the new group of projects, although in the majority M&E still is uncoded. If there are M&E elements they are usually incorporated in the project management and technical assistance items. One recent journal article on M&E in rural development found that the average cost of M&E as a percentage of total project costs was between 2 percent and 3 percent, though for a few projects the percentage was much higher.²⁴ Those are small numbers, and may suggest that the Bank has been systematically under-financing a component that required more resources. On the other hand, trying to assign an indicative upper limit on M&E as a percentage of total costs runs a big risk. To be effective, M&E

24. "Assessing M&E - Has Project Monitoring and Evaluation Worked?" Maddock, Nicholas. In Project Appraisal; Vol. 8, No. 3, Pages 188-192. September 1993.

requires whatever funds it takes to inform management - and the financing agencies - about project progress and problems. In most projects the budget demands will be small. In the more experimental projects, a budgetary ceiling on M&E may disable the effort to learn lessons and respond with the right adjustments.

B. Agriculture

(i) The Golden Years

4.8 Subsequent sections describe a collection of M&E experiences which cannot escape the label "disappointing". The word is appropriate for the sample of projects that was drawn for the study, mostly of operations that covered the middle and end of the 1980s. Some good M&E performances were missed even in that period. For example, the successful teaching and subsequent networking of M&E principles and practices to local staff in the West African countries that borrowed for agricultural extension projects. This group of M&E country enthusiasts has maintained its momentum, and a series of joint regional seminars on extension M&E, despite the transfer of the Bank's regional M&E advisor who brought the group together. None of these West African projects were picked up in the sample. Nevertheless, the study design ensured that the sample represented a large majority of all projects in operation in the middle and late 1980s. Thus, these examples of good performance must be considered exceptions.

4.9 A more important loss from the Overview is the set of rural development and irrigation projects which were brought under the influence of AGRME starting a decade before, in the mid-1970s. Para 3.2 pointed out that the golden years of M&E predated the sample, and that already by the middle 1980s M&E as well as KPI were both on the wane. When AGRME was in its prime, however - and that coincides with the period when the rural development program itself was running at full speed and enjoyed Bank-wide respect - the impact of AGRME staff and their consultants (and sympathetic colleagues in AGR's Rural Development Division and the regional agricultural divisions) was substantial on the projects they were visiting.

4.10 The spirited intervention reflected the group's conviction that progress in rural development was not "linear", and conventional monitoring practices and indicators were inappropriate to guide the evolution of this important portfolio. M&E was a point of entry not merely to measure and evaluate results but to turn the interests of project managers to social and institutional issues that were of small concern to the bulk of the Bank's portfolio but critical to make this new generation of ambitious projects work. They were able to attract in several countries equally enthusiastic teams of consultants, often university-based, which were also dedicated to push the operations towards their objectives. The M&E intervention from these external sources was strong, and the impact on projects, impressive. Box 7 describes the M&E activity in preparing for and putting in motion the Northwest Region Development Program in Brazil starting in the late 1970s. This is an example of good M&E at work - relying on strong external support for the "E".

4.11 These years of exceptional M&E activity, strongly supported by Bank and consultant staff, tell a different story than the one that follows. They show that project M&E, if well handled and supported, can reach beyond its immediate objectives and influence the course of the entire project operation. The next paragraph shifts to a different scene, where issues of ownership and sustainability come to the fore. The two periods offer important, contrasting lessons: what M&E offers when the Bank takes a "hands-on" role in putting it to work, and what has usually happened when the Bank (or any donor) lets go. It is the lessons from *that* contrast that are the most important. Given the Bank's interest in host-country participation, and in encouraging the spread of the learning culture, a return to the earlier hands-on approach is no longer tenable (as a general prescription).

Box 7: M&E in the Brazil Northwest Region Development Program

Background The constructive role which can be played by M&E at different stages of the project cycle was demonstrated in the case of POLONOROESTE, a program which dealt with many unknown variables, in a vast frontier region and a complex, new institutional setting. POLONOROESTE consisted of a package of initially 3, later 5 loans for the states of Mato Grosso and Rondonia, for upgrading a central regional highway; agricultural and infrastructure development in frontier settlement areas; environmental and amerindian protection; health and education. Most of the program area was under primary tropical rain forest cover when the program was initiated in 1982. The basic objectives of the Bank loans were all associated with orderly frontier development: consolidating the economic and social conditions of frontier settlements, fostering prudent environmental management by insisting on a restriction of forest clearings and agricultural settlements to exceptionally fertile soils; protecting and managing forests, biological reserves, national parks and the numerous indigenous groups (some of them never contacted) and their reserves. M&E received systematic attention during project appraisal. It was - together with the negotiation of the very first "Amerindian Protection" component - the most contentious issue during the lengthy negotiations for these loans. Ultimately, monitoring was assigned to project management, to support day-to-day implementation and decision making, while evaluation was to be contracted with a major university, which set up a multi-disciplinary team for this purpose. AGRME staff made frequent visits during the first half of the program.

M&E Achievements In Project design: the careful consideration of meaningful M&E indicators, and of the potential usefulness of M&E-generated information, disciplined planning and added clarity. **In the Start-up phase:** M&E - by establishing what was to be achieved - provided a focus for training of recently recruited project staff and for sensitizing the institutions involved to the novel objectives and guiding principles of the program. **In early implementation:** it readily brought out shortcomings, by obliging the consideration of qualitative questions which went beyond the customary reporting on disbursements and quantitative measures of delivery of project inputs. It brought with it a systematic "listening to the people", participation and participant observation, as the ongoing evaluation team spent several weeks in the Region every year - the anthropologists and medical staff in indigenous communities, and the agriculture and economic experts in the areas of rural settlements. In addition, satellite imagery provided objective pictures of the unprecedented scale and pace of forest destruction which ensued as an unexpected, indirect outcome of this program, especially construction of the roads. **At mid-term (after the first 2 years):** the evaluation system, in conjunction with some very fine field level, community focused monitoring, brought out and documented all that was not well with this program. For example, the violation of environmental and amerindian covenants, the unsustainable agricultural practices and incipient land abandonment, and the inconsistent, if not fraudulent implementation of health programs by local authorities. It documented the virtual derailment of some of the program components, as the Region was flooded with migrants, poor peasants as well as land and timber speculators and gold and diamond miners. Without the obligation to jointly carry out a mid-term review, the Borrower agencies and the Bank would have taken longer to establish the seriousness of problems and they would have faced higher political obstacles to make their case heard. **As a corrective instrument:** Mid-term evaluation results provided the basis for a de facto suspension of disbursements, and for the design of a detailed Action Program. They laid the ground for longer term design adjustments, incorporating the lessons of the early 1980s. They also made for a lively and often acrimonious local and country-wide debate and - ultimately - constructive, corrective strategies and a certain measure of consolidation.

Caveats POLONOROESTE demonstrates that, when M&E is to the point, it is also bound to run into a number of problems. The culture of constant critical review and adjustments can be at variance with the general administrative culture. There will be some who consider it an improvement to discover and correct shortcomings. There will be those who find it threatening and disturbing. The problem looms larger when it comes to periodic, independent outside evaluation. Project management is the main client, but only as a recipient of a product, not as a controlling agent of the evaluation process. In POLONOROESTE, project management at times tried all its tools - vetoing access to information, claiming that evaluations were not carried out to the letter of the contract or competently, or not releasing travel funds or salaries. Much was done to make sure the process was as difficult as possible. Built-in tension between a top-down government apparatus in charge of project management and an intellectually and grass-roots inclined outside evaluation agency can deteriorate into temporary tug-of-war scenes where mutual accusations of improprieties abound. In designing M&E, checks and balances must be built in to mediate such conflicts.

(ii) The Sample

4.12 The record of the 21 old agriculture projects in the sample provided little information on the success of the M&E components described in the SARs. One is dealing here with PCRs, and a few PARs, that did not say much about the success or failure of any parts of the M&E component. Conclusions based on this record most be treated with caution. Nevertheless, when M&E and KPI are mentioned in the PCRs, the results are usually described as inadequate and disappointing. That extends also to the only two, high-profile M&E cases identified in the old set - Bangladesh Rural Development II and Thailand Land Titling - where substantial expatriate M&E technical expertise was made available but could not produce the expected results. Usually the pattern is less clear. The SARs for this old agriculture set were generally not M&E-intensive. The ex-post reports say that in most cases even those low expectations went unfulfilled, or that the M&E component may have lived up to expectations but the results in terms of impact on management were what was disappointing, i.e. little M&E output or not enough to have provided the data to management and the Bank to make the informed decisions that M&E is supposed to support.

4.13 Since most of the authors of the PCRs and PARs were themselves untrained in detecting and assessing M&E, it is possible that the actions and results of M&E have been systematically discounted in the documented record. This point is illustrated by the comment of a Bank task manager working on an Asian agricultural credit project. When asked about the lessons he learned from the credit agency's evaluation reports he said he did not know and did not care - that they were all junk and he sent them unopened to the archives. It is possible also that some successful projects with no announced M&E at either the SAR or PCR stage had functioning, effective and quiet information and monitoring procedures. These options were expected to be clarified in the Bank-wide interview program that followed the desk review. But, as mentioned already, the interviews failed to clarify this point - usually because the staff could not recall anything about what they did not pay attention to earlier.

4.14 Among the documented project histories where M&E *was* written up, some are truly depressing:

- in Indonesia an agreed M&E input to a rural development program was not delivered till near the end of the project life, and the PAR claims that in the absence of M&E the project continued to promote an unprofitable soil conservation scheme that was abandoned at project completion;
- in Tunisia the SAR implementation annex that detailed the M&E and KPI appraisal plans of another rural development project was never translated and sent to government, and the consultant proposal that was eventually presented was too sophisticated and inoperative;
- in Zimbabwe the expatriate cohort that remained in the ministry's extension service was unable to activate the full appraisal M&E design for an agricultural "services" project, for reasons that are not provided in the PCR; etc.

4.15 The overall picture painted in these PCRs is of M&E that is not coming anywhere near its potential and of Bank supervision that is unable or unconcerned to do anything about it. Whether all the apparent poor performers in the sample are genuine, or the reflection of inaccurate reporting, are questions that the study could not pursue in this group of 21 old agricultural projects.

Nevertheless the sheer weight of negative commentary in the PCRs and interviews alike suggests that these cases of misidentification must be the exceptions.

4.16 Within this depressing panorama there are a few positive stories of overachievers: the Hungary Crop Production Improvement Project and China Forestry Development Project are two examples where the M&E (and MIS) contributions were actually much stronger than the SAR had proposed (these reversals explain the improvement in the overall subset noted in para 4.3). The China case is an example of a much broader characteristic of that country portfolio which distinguishes it from the rest of the world. The Forestry Ministry is one of the best among a majority of Chinese official agencies with a strong commitment to data collection. The Bank did not have to emphasize or even request at appraisal that the data continue to flow. Box 8 elaborates on this story, and discusses the limitations on this massive data bank and the way the Chinese use it.

Box 8: Data Collection and Use in China

Bank efforts to implant project level M&E in its China portfolio have to adjust to circumstances rather different than in other borrowing countries - if not unique. Normally the Bank pushes governments to create and expand the data bases for monitoring and subsequent analysis. In China, however, there is no lack of basic data. Historically Chinese governments have been massive consumers of information from provinces and public agencies. This reporting "culture" is partly explained by the contractual nature of tasks: operational units at all administrative levels are obliged to report progress figures to superior levels to support the financial contract. (Bank staff are familiar with the habit of Chinese officials to begin discussions (and banquet speeches) by quoting figures of project results.) Data collection is so routine as to need no prompting, except when new-style projects reach outside the normal flow.

The problems for the Bank in promoting M&E in China lie elsewhere. One is that the Chinese can be protective of their information, anticipating that it may be construed or misinterpreted by others and have the effect of discrediting them. The Chinese, at all levels of project administration, route Bank staff to higher levels of authority for permission for access to the data bases. As a part of this scenario, the Bank occasionally is provided falsified or misleading statistics, especially prepared for donor agency consumption. But this is not common. There are as many cases of the opposite result - when the data present a view of project activities which is positive, robust and unembarrassing. The Chinese Forestry Development Project (1985-1992), one of those included in OED's study sample, is a good example. This was an activity where the executing agency - the Forestry Ministry - not only has a massive data retrieval system but is willing to share it with the Bank.

The other problem is a cultural phenomenon. The Chinese collect masses of data, but use it mostly to draw simple averages to describe current conditions and trends. They do not routinely approach these files as a base for probing for relationships and lessons. They are a monitoring, not an evaluating, culture. They rarely do regressions or anything other than simple statistical analyses. Thus the Bank had to persuade the Forestry Ministry to exploit its forest resource data base in new ways, to determine, for example, whether operational problems reported from field stations in the provinces could be associated with certain soils, species, planting practices, etc. In the second forestry project presently being implemented (National Afforestation Project), the Bank persuaded the Ministry also to start collecting price information on sales of the different species, to match with the ample information already available on sales volumes, to determine whether the planting programs for the next season ought to be revised to reflect changes in consumer preferences. The Bank can assume that a more analytical approach to the data bases will emerge. But at this time the Bank's efforts to promote M&E in China have more room to expand in E than M.

4.17 The academic literature in recent years has featured several good articles on M&E. A well written and useful synthesis of current concerns published in 1992 - by a veteran of the Nigeria M&E experience (see para 4.25) and currently a lecturer at the University of East Anglia, is repeated in

this report as Annex 2.²⁵ He discusses at length many of the factors that are widely agreed to be limiting the progress of M&E programs, most of which are picked up in the discussion of the case studies in the next section.²⁶ Among other evidence, the author cites OED's own 1988 report on rural development, which noted that of 104 Bank projects with built-in M&E components "only 15 percent showed good M&E results, 39 percent had seriously deficient M&E systems and in 46 percent the M&E system was either not implemented or performance was unsatisfactory". By any standards, the satisfactory rates were astonishing low.

(iii) The Case Studies

4.18 Searching for evidence that could help explain this record of poor performance of M&E in agriculture during most of the 1980s, OED conducted a series of rapid case studies of seven agricultural programs supported by the Bank - all with a reputation for high M&E profiles. These programs lead the ranking of Bank inputs into M&E, and one would have to have assumed a priori that their group serves as a good measure of overall M&E performance in the sector (and notwithstanding the impressive exceptions - see paras 4.8 and 4.16). Visits were made to five of the countries. Two African cases were taken from information supplied by consultants on other assignments. Summary statements on each of the cases are provided next, followed by general comments on the implications of the group. OED is preparing reports on each case study, which will be collected and available as a Working Paper.

4.19 **India: T&V Agricultural Extension Projects.** Two state extension projects (Madhya Pradesh and Andhra Pradesh) were audited during the study period, paying particular attention to the M&E components. Another five had been audited by OED in 1990. OED also financed a consultant report, for the study, reviewing the factors that determined the relative success of M&E across a sample of eight states with varying M&E performance. The objective was to find some answers applicable to all 17 of the states of India with T&V/M&E programs. A summary of the consultant's report is attached as Annex 3, the highlights are presented in Box 9, and the full report is available at OED as a Working Paper. Though a few of the state M&E programs are considered modestly successful, the majority are not. The common criticisms are that they have not been properly staffed and have slavishly followed the reporting requirements proposed by the Bank in the early 1980s, without adjusting to the evolving interests of senior departmental officers. The level of data quality, analysis and interpretation is low and the results have rarely been used by state managers in the extension service or department of agriculture. Success is seen to be linked to the support of strong individuals of high rank. The M&E Unit in the Extension Directorate in the Central Government is understaffed and unable to secure significant improvements in the system. The Bank has made no technical input into M&E operations since the Bank's M&E advisor was reassigned in 1987.

4.20 **Pakistan: Drainage and Irrigation Projects.** OED examined M&E socio-economic systems of projects implemented since 1982. The earlier SCARP deep tubewell programs had no socio-economic M&E. Governments at the national and provincial levels have given little support to Bank attempts during the 1980s to incorporate socio-economic evaluation along with the monitoring of "physical" phenomena, e.g. the water table, salinity level, etc. Consultants have produced evaluation reports, but these have largely been ignored by government *and* the Bank. The physical monitoring was also inadequate. This is even the case with two pilot programs where, at appraisal, the Bank considered that monitoring was most important: for the so-called "transition" projects, which were aimed at shifting tubewell ownership to the private sector, and for the "tile drainage" projects,

25. "Monitoring and Evaluation in Agricultural and Rural Development Projects: Lessons and Learning." Coleman, Gilroy. In *Journal of International Development*; Vol.4, No.5. Pages 497-510 (1992).

26. Footnote 24 to para 4.7 cites another recent journal article discussing the limitations on M&E.

introducing a new system of horizontal drainage. One unusual feature of both series of these Pakistan pilot projects is that the M&E was substantially under-funded in relation to the M&E objectives. This, coupled with long delays in executing the investment programs, has meant that the projects' M&E components have run out of Bank funds before the investments have produced any results. In effect, the socio-economic surveys and studies are still updating the baseline. The Bank hopes to find more funds, but it needs support from governments that are still unconvinced as to the value of the studies.

Box 9: M&E in the Indian T&V Agricultural Extension Program

The Indian consultants to OED (the Centre for Agricultural and Rural Development Studies - CARDS), selected for intensive interview six of the seventeen states which received Bank support for T&V extension. CARDS developed a list of criteria for measuring, rating and comparing the quality of the M&E unit's performance in each state extension program, and a separate list of criteria for assessing four sets of factors which were expected to influence that performance. These sets were: organizational aspects (including location of the unit), personnel aspects (including staff strength, skills and tenure), support facilities (including budgets), and the Unit's "interaction" with users (including rapport, interest and involvement of senior ministry staff).

Written responses to the formal questionnaires were not as productive as follow-up interviews conducted by the two CARDS' study directors. The issues and lessons that came to the fore are described in detail in the consultant's report, of which the summary is annexed to this paper. Of the six state T&V programs, two were classified as high, two as average, and two as low M&E performers. CARDS' job was to try to explain those differences by the four sets of factors or any other observable patterns - apart from idiosyncratic and largely unpredictable behavior. Two other, economically progressive states were added using rapid assessment in order to enhance the study base - Tamil Nadu, where M&E outperformed all the other states, and Punjab, which remains largely indifferent to both T&V and T&V's M&E, reluctantly accepted a T&V project, but refused to set up a separate M&E unit.

The report discusses the rise and gradual decline of T&V as well as M&E operations since the Indian T&V program was established in the late 1970s. T&V is still considered a superior system to the extension programs which preceded the Bank's intervention, but the early optimism for rapid and continuing production impact has faded. However, that appears to explain only a part of the general lapse in support for M&E in six of the eight states. The report mentions the inflexible survey program, which still tracks routine indicators peculiar to T&V (number of contacts, number of messages received, etc.) that were prescribed by Bank staff in the early 1980s and in which at present "the users of M&E findings have virtually no interest" (page 23). It mentions the hierarchy-oriented administrative structure in India, where "any unit purporting to report on the achievements of another unit is bound to invite suspicion" (page 21), and it mentions the failure in most states for the unit to be internalized in the ministerial culture, so that they depend for their survival on promises of further Bank support (page 98). CARDS examined the hypothesis that M&E received greater support from Directors of Agriculture who were appointed from the technical streams, compared with those appointed from the cadre of the Indian Administrative Service (IAS). Incentives, interests and criteria for promotion would differ substantially between these two streams. The consultants were unable to arrive at a clear judgement on this issue.

With respect to the four sets of explanatory factors specifically targeted by the study design, CARDS found that the quality of the interaction with users (a dimension of "ownership"), and the personnel aspects appeared to be critical. The organizational and support factors were not. CARDS speculates that a strategic position in the organization, and an adequate budget, were necessary for good M&E, but that these characteristics were shared by the poor performers as well. In these respects, the Indian T&V case study arrives at the same conclusions as OED in the M&E Overview. CARDS did not inquire about the level of Bank support and ownership of the M&E programs, which was the third major factor identified by OED as necessary for good M&E (see para 4.29 (2) of the Overview).

4.21 Indonesia: Irrigation Investment and O&M. Substantial M&E survey capability has been built into the Bank's irrigation projects implemented since the early 1980s. M&E was insignificant in previous projects in this twenty-five year-old series. The Bank and government have been satisfied with the results of the investments in new and rehabilitated infrastructure, and with the newer series of projects dedicated to operations and maintenance (O&M). Weakness in water management at the tertiary canal and on-farm levels led the Bank to strengthen the M&E side. This has largely been organized and executed within the expatriate-supported consultancies that have implemented the schemes. Government and local ownership, and sustainability, of the surveys and special studies is uncertain. The field data is collected by regular provincial staff, and sent through provincial and national headquarters to the Bank without much scrutiny or analysis. One analytical ("statistical") procedure developed by the Bank and the consultants, and carried out faithfully by the Indonesian staff without full understanding, has recently been shown to be misapplied and misleading. The Bank's agricultural division was disillusioned with progress in M&E under the first of the O&M sector-level projects, and reduced Bank support under the ongoing second project. Thus, despite technical assistance in the 1980s by the Bank's regional M&E advisor, ownership by the Bank as well as government is in question and supervision staff on both sides are not trained to make judgements on the formal analytical procedures in use.

4.22 Mexico: FIRA's Agricultural Credit Program. In 1976, during the Fifth Project, the central rediscount agency "Trust Funds for Agriculture" (FIRA) reluctantly acceded to the Bank's request that FIRA establish a systematic M&E capacity. The idea was to measure the impact of FIRA's credit operations. Although further efforts were made to upgrade the M&E system during the Seventh Project, the Bank's interest later declined. Indeed, FIRA complained to OED staff that when plans for upgrading the system were sent to Bank Headquarters for comment they attracted no response and that supervision missions no longer provided the guidance that FIRA needed. Nevertheless, FIRA set up a computerized M&E system that tracks financial and production data for a stratified sample of about 700 farms across the country. The last two PCRs on the FIRA series of projects pay practically no attention to this home-grown monitoring and evaluation effort. The lack of interest in FIRA's M&E was not confined to the Bank. The PCR for the Seventh Project notes that "monitoring and evaluation had suffered from the lack of support from [FIRA's] senior management and staff, and it was perceived for a long time as a burden imposed by the Bank rather than as an important tool for impact analysis of credit on Mexican agriculture". This attitude has persisted: the evaluation system is rarely consulted by management or by field level staff and was nearly abolished in 1991. However, a system of analysis of profitability of various lines of production, which was introduced in 1991, has been well received by FIRA's staff. FIRA currently has plans to integrate it with the evaluation system and with two other databases on output and input prices in order to establish a unified system of greater relevance to FIRA's needs.

4.23 Brazil: Northeast Rural Development Programs. The Bank and its counterparts in the federal and state governments who helped design the POLONORDESTE and successor PAPP rural development programs intended to create a major M&E capability at the state and regional (SUDENE²⁷) levels. During most of the first decade of POLONORDESTE - starting in 1975 - the Bank gave strong support to the M&E operations. Then the support and the operations declined. During the PAPP period - starting in 1985 - the Bank gave intermittent technical support to get the M&E program back on track. But even these efforts were short-lived. The inconstancy of Bank support to M&E in Northeast Brazil is described in Box 10. The M&E units within the state project coordinating units and in SUDENE suffered throughout the late 1980s from poor staff appointments

27. The federal agency, headquartered in Recife, created in 1959 to promote development in the northeast.

and under-funding. When the Bank dictated two successive reformulations of PAPP design in 1991 and 1993, the previous M&E program should have been reoriented. Instead, SUDENE and all five states visited by OED during the case study disbanded their evaluation staffs and abolished the units. SUDENE and the states are presently depending on the Bank's Recife Office to prepare and disseminate new guidelines for both monitoring and evaluation. This Brazilian experience in the northeast during the late 1980s stands in marked contrast to the M&E experience in the northwest during the first half of the decade - described in Box 7.

4.24 Malawi: Area and National Rural Development Programs. Starting in 1968, this was the first of the big rural M&E programs in the Bank. It was well funded, initially with massive expatriate support. Criticism about all the Bank's rural M&E programs started here - about their failure to clean and use the large and dirty data files that were being produced from repetitive field surveys by undertrained enumerators. The national M&E programs continue, with lower funding from national sources, but the image of poor performance persists. Staff appointments have not brought in well-trained analytical strength, and rotation levels are high. A few expatriates continue to play essential roles in keeping the program going.

4.25 Nigeria: Agricultural Development Programs. Beginning in 1975 with the establishment of the federal M&E unit in Kaduna - APMEPU - this was the flagship of the Bank's M&E programs. The number of expatriates in management, evaluation, and statistical assignments in APMEPU and the individual state ADP M&E units reached 26 in 1982. The number declined precipitously thereafter, and the foreign talent was never replaced by comparably trained local staff. Yet despite the expatriates, APMEPU and the ADPs produced a data base that was to be later criticized as dirty and unusable for measuring progress towards the primary yield targets (as in Malawi). The Bank sent a mid-term review mission to APMEPU and the states in 1985. The mission considered the situation desperate but not irretrievable. Progress since then in terms of cleaning the files, strengthening survey validation and reporting procedures, and appointing higher quality staff, has been minimal²⁸. APMEPU has been renamed (APMEU), reduced in size, and suffers continuous under-funding. APMEU is supervised from the Bank's Lagos office, and a recent mission included consultant support to try to plan a recovery.

4.26 For all seven programs, the Bank at appraisal intended to establish a strong M&E capability. The SARs in some cases did not explicitly identify KPIs: rather the emphasis everywhere was on establishing units within the executing agencies adequately staffed and funded to carry out field surveys and special studies of both implementation performance and project impact. Indeed this group of programs distinguishes itself from both old and new programs in other sectors by the intensity of agency-level M&E activity and the attention given to measuring impact on farm enterprise and beneficiary incomes. Though the M&E inputs by the Bank varied from project to project in each program, taken as a series the resources for M&E put into the SARs of each of these programs would be ranked at least as high as the best M&E projects approved in 1993. They would not have included participatory evaluation and some of the other features that distinguish the new set, but they would have far outdistanced them in attention to building M&E capacity in the agencies. Though the comparison with the new set of high profile M&E projects is not exact, this group of seven projects provides important lessons about what can happen during implementation to an M&E program established in strength in appraisal documentation.

28. APMEPU's actions in response to the 1985 mid-term review recommendations is discussed in the collection of working papers being prepared on these case studies (see para 4.18 of the text).

Box 10: M&E in the Brazil Northeast Rural Development Program

The Bank has supported two generations of rural development projects in Northeast Brazil - POLONORDESTE (1975-1989) and PAPP (1985-present). Their objective has been to improve production and welfare in this drought-prone area comprising nine states and part of a tenth, about 30% of the country's population, and income levels half that of the rest of the country. PAPP replaced the earlier program, and dropped most of the components that were not related to agriculture. Both programs have been coordinated by SUDENE, the federal authority established in Recife in 1959 to support development of this disadvantaged region. M&E was an integral and vigorous element of POLONORDESTE, and the enthusiasm and contributions of Bank staff and consultants to the M&E operation in the northeast was a prelude to similar, early successes in the subsequent program for the northwest (Box 7). However, in 1983 OED made a special review of this experience (see footnote 3). It concluded that the M&E operation was less effective than commonly reported, and that the Bank should provide more support. PAPP appraisal reports maintained the attention to M&E. But in practice Bank intellectual and technical inputs into the M&E systems in the states and SUDENE have declined even further.

PAPP technical assistance funds for the SUDENE component financed consultant visits in 1985 and 1986 to redesign the M&E program to suit the PAPP objectives. SUDENE prepared new guidelines and sent them to the Bank for comment and approval. The Bank did not respond. Under the uninspired leadership of the then head of the "Evaluation Unit", SUDENE let the M&E work drift and the state survey activities to decline. One reason for the poor performance was the process for appointments in SUDENE, which staffed the Unit with persons, released from other divisions, that it was having difficulty placing. In 1989 the Bank's Recife Office attempted to rejuvenate the program. A British firm specializing in M&E^{a/} was contracted to design a survey program consisting of two special studies and a pilot round of a more general farm impact survey at the state level. The intention was to get quick, significant and operable lessons to reestablish the credibility of M&E with management at SUDENE and the state PAPP coordinating units. Following the design work, the firm provided a short-term (nine month) resource person to help guide the Unit and, together with Unit staff, visit the states selected for the surveys and get the pilot studies and general survey moving. In 1990 the first reports were ready and a seminar brought the state and SUDENE staff together to discuss future plans. The input of the expatriate was appropriate, and appreciated by SUDENE and state staff.

The next year the program disintegrated. By 1994 the evaluation units at SUDENE and each of the five states visited by OED on a special M&E mission that year had been abolished ("extinct" in Portuguese). One of the reasons was that the Recife Office had not been able to persuade Bank Headquarters to take the initiative seriously. The Headquarters back-stopping officer had previous experience with project M&E in agriculture, and was convinced that the consultant-led effort would fail. No more funds were put into the TA contract, and the firm left the field. Another reason was that the Bank had come to the conclusion that PAPP itself had to be redesigned, in part to deal with the unanticipated, precipitous decline in federal funding. After two such "reformulations", the 1989 M&E program no longer made sense. SUDENE and the states have depended on the Recife Office to provide new guidelines for both M and E. Bank Headquarters is no longer involved with the M&E program of PAPP.^{b/}

^{a/} Information Technology and Agricultural Development (ITAD), the only consulting firm which has specialized in M&E.

^{b/} Independently of these project M&E activities for PAPP, OED sponsored in 1988 a consultant to evaluate those elements of POLONORDESTE and PAPP which appeared to work well. Her report is one of OED's best sellers ("New Lessons from Old Projects: the Workings of Rural Development in Northeast Brazil". A World Bank Operations Evaluation Study. Tendler, Judith. 1993.)

4.27 To organize the lessons from these case studies, OED identified eight hypotheses about the common characteristics of M&E plans that are successfully implemented. Since the M&E components in all of the seven programs performed substantially less well than expected, the purpose here is to determine whether one or more of these factors was consistently missing in all cases:

- (1) That *ownership by government*, or at least by project management, had to be present at start-up or develop over time. Among other things, this means that government and project management had to have a vision of what M&E would provide to them, and not simply accept the Bank's claim that the operation was useful.
- (2) That the *Bank* also had to demonstrate a sense of ownership for the M&E program, extending beyond the appraisal period and the departure of staff responsible for initially designing the M&E program. Continuity is presumably best guaranteed by support from within the sector division, although specialized M&E inputs from technical support divisions are an option.
- (3) That the quality of the *appointments of nationals* to senior, technical and analytical M&E positions was adequate, that good performance and experience of the M&E nationals was rewarded, and that turnover rates were modest.
- (4) That the M&E unit was provided a *budget* adequate to support the planned field surveys and special studies.
- (5) That the projects themselves had to be relatively *successful* in achieving their goals, or, in the absence of impact data, at least be perceived to be successful in meeting implementation schedules. Otherwise, a defensive reaction would set in and the interest in self-examination would fade.
- (6) That the M&E program did not depend upon long-term *consultant support*, or at least that responsibility was successfully transferred to nationals over time so that their sense of ownership was cultivated rather than frustrated.
- (7) That the M&E plan should not be *overambitious*, recognizing that, even with good local staff appointments, overloading the unit would risk dilution and poor results across the board. The primary enemy here, cited in all reports speculating on the reasons for success and failure of M&E (see Annex 2), was the popular application of large scale, repeating farmer field surveys trying to measure production impact and, in the early years of M&E, labor inputs and household income.
- (8) That the *organizational location* of the M&E activity mattered. Without anticipating the preferred position, one can speculate that placing the activity *inside* the project close to the project manager, as well as keeping it *outside* the project, or at least outside the reach of a manipulative manager, both have advantages. M&E professionals tend to favor a split in the system, with monitoring inside and evaluation outside the project. The seven projects were defined in terms of where both M and E were located.

4.28 The literature on M&E mentions many additional factors that influence the success of the program (see Annex 2 for a good summary). Examples are the timeliness with which results are put into the hands of management, excessive emphasis on attempts to evaluate impact at the expense of monitoring implementation, emphasis also on "hard, production" data, rather than "soft, attitudinal"

changes, and credibility of the data base. The case studies did not provide information to permit investigation of these and other plausible arguments.

4.29 The three factors which appear to be deficient in most if not all cases are:

- (1) Ownership by government, and, in particular, project management. The special consultant report on India for OED (Annex 3) shows that relative success of T&V extension is closely associated with the rapport established and support provided by leadership in the agricultural ministries. In most states that support was lacking, and the contribution of M&E to policy has been minimal. Ownership of M&E in the Brazilian PAPP program is non-existent (refer to para 3.11 for a contrast with some other Brazilian projects). It is hard to establish that government or project management had a "vision" for M&E in any of these cases (other than the M&E staff itself). They were mainly following the Bank's lead. This is not a favorable situation (though admittedly it avoids the other, potential problem of the Bank and government having conflicting visions).
- (2) Continuity of attention by the Bank was deficient in all cases as well. That the Bank had a "vision" is also suspect. There was no sustained Bank support for M&E after the initial start-up period. The typical pattern was for the Bank to rely on expatriate project consultants to set the survey's initial design and get the programs running. Bank supervision was mostly interested in picking up the survey results and bringing them back to Headquarters.²⁹ Whether due to lack of training, interest, or priority, supervision staff did not actively promote the consolidation of the M&E program and rarely reported on its performance. The often skeptical attitude of Bank staff toward substantial M&E components planned for agricultural projects under their control is reflected in two mini-histories of frustrated M&E in Ethiopia and Zaire presented in Box 11.

There are exceptions: the staff member appointed by the South Asia Region to assist the development of M&E in those programs visited India and Pakistan for several years after project effectiveness. When he transferred, the Region's support stopped. The same cycle was repeated for the M&E expert appointed to the Africa Region, and the links she forged between the country M&E units have continued even in her absence (para 4.8). The East Asia Region offered similar technical back-stopping in Indonesia (as well as China, Thailand and the Philippines). But in this instance the responsible agricultural division has disputed the effectiveness of the M&E program of a recently completed project, refused to carry it forward into the successor project, and left it to the Indonesian's to prove that the ongoing M&E system was worth further Bank lending.

- (3) Appointment and retention of qualified nationals to the M&E unit was deficient in all cases.

29. A British contract officer in charge of the Evaluation Unit in the Bauchi State ADP in Nigeria (para 4.25) lamented that Bank supervision staff had never visited his unit's offices to observe the operation and his computers; that meetings for him were invariably at the project manager's office. Bank staff did visit agriculture and other divisions. This lament has broader significance than as evidence of Bank inattention. The fact that the Evaluation Unit was located at driving distance from the project manager's office reflects as well on the manager's disinterest. He paid no more attention to the evaluation operation than the Bank did. On the other hand he set up a separate Monitoring Unit, and put it and its personal computer next to his office. That was the information he wanted: on the progress of implementation.

Box 11: Examples of Bank Staff Skepticism about Substantial M&E

Bank project task managers are in a position to promote or discourage M&E initiatives. The latter response arises from genuine concern that M&E benefits are overstated, M&E skills are uncommon, and a "substantial" M&E operation has a low pay-off and wastes Bank and country resources. Two examples from Bank experience in Africa in the 1980s illustrate this skepticism, and its impact on the processing of M&E designs. In both these cases, other factors were to intervene to undermine the effectiveness of M&E regardless of the level of committed resources. In both these cases as well, a difference of opinion emerged in the initial years - when M&E may still have had a chance - between the Bank and IFAD (cofinanciers) about the potential of a substantial M&E system. But that conflict is uncommon in the history of collaboration on M&E between these two agencies, and is **not** the point made here.

In Ethiopia, the Bank restarted its lending program in the late 1970s by funding the second phase of the integrated rural development program called the Minimum Package Program - MPP II. IFAD proposed to finance a separate project for expanding seed production, but this was later incorporated in MPP II and the two agencies cofinanced the enlarged operation. The Bank wanted a modest M&E program, and helped recruit an expatriate M&E advisor for the Ministry of Agriculture. However, a senior member of IFAD's Monitoring and Evaluation Division visited Ethiopia - independently of any Bank mission - and subsequently prepared a proposal for much larger M&E activity. The Bank's regional mission in Nairobi, which had primary responsibility for supervising the project, received a copy of the IFAD proposal. Bank support was essential to put the new plan into effect. But the Bank's task manager felt the proposal was excessive, and of lower priority than other urgent operational problems. He literally "shelved" it. It was rediscovered on the shelf years later, and discarded. This use by the Bank of its controlling position proved to be immaterial. After the expatriate had produced his first report of MPP II's problems as well as progress, his transport was restricted and he was later reassigned to assembling quarterly reports and preparing for a possible successor project. The Ministry was not interested in either a small or large M&E operation, if it attracted criticism and attention.

The second example started in the early 1980s in Zaire, and again featured a Bank task manager and the same IFAD officer. In this case the task manager had inserted an M&E component into appraisal design of the Smallholder Maize Project - also planned for cofinancing by these two agencies. But he allowed no donor finance for the M&E Unit, setting it up to depend exclusively on host-country contributions. This effectively doomed full execution of the component, a fact recognized by the Bank officer when he planned it and by IFAD only after the Credit was signed. IFAD tried for several years to have that part of the agreement corrected, to allow for external resources in support of M&E. The Bank's task manager remained reluctant to accept this revision, since he was convinced any extensive M&E activity would be a waste of donor and counterpart funds - given the absence of base-line information and the requisite staff to make M&E effective. IFAD eventually succeeded in redirecting some funds, but the project as well as the M&E action were equally unsuccessful and the operation was considered a failure after a few good, first years. The Bank's task manager argued then and argues now that the outcome of the Bank-IFAD standoff did not matter to the fortunes of the project. IFAD's position was that better M&E was within reach and might have helped management move closer to project objectives. There is no answer to that debate. Neither party would contest that when the conditions for effective M&E can be established, it can demonstrate its utility.

4.30 No consistent pattern emerges for the other five factors, at least from the limited data available to this review:

- (4) The budget factor does not help explain the overall outcome. In four countries the M&E budget was clearly under-funded, but only in Brazil is there evidence that the provisions were so low as to emasculate the planned survey programs. In two other countries (Indonesia and Nigeria) budgets were not restricted and M&E results were nevertheless disappointing.

Adequate budgets are a necessary condition of good performance, but the budgets provided in this group of programs were enough to get better M&E results.

- (5) Most of the programs supported by these M&E activities have not delivered anywhere near the level of economic results anticipated. Three of them are rural development programs, a class of integrated management activities which the Bank now believes to be inherently flawed. Nevertheless, the key point here is whether the project was perceived to be on track, and from that perspective two of the three rural development programs (Malawi and Nigeria) were considered by government to be effective applications of donor funds. The governments of Indonesia and Pakistan are equally convinced that the irrigation programs supported by the Bank are important and successful applications of public funds. Mexico is satisfied that FIRA's credit lines provide essential liquidity to finance farm and crop development. Finally, the T&V extension system in India is seen to be falling short of its promise, but again the perception in the Government of India as well as the majority of states is that T&V is nevertheless superior to the rural extension system which it replaced. PAPP in Brazil is the only program which attracted very little praise during the majority of its life to date. Thus poor performance of M&E for the set of seven projects as a group cannot be associated with disappointing results for the projects.
- (6) The consultant factor plays unevenly across the group. India embraced the T&V methodology quickly, and required no expatriate support to put it into practice (apart from the visits and encouragement of the person who created the T&V system, and the Bank staff member mentioned in point (2) in the preceding paragraph). In Northeast Brazil, consultants played a minor role, there was no problem of transfer of responsibility, and the M&E program collapsed anyway. More, not less, consultant support was needed. In the other four countries, however, the hypothesis that transfer to locals is essential is confirmed. There, consultants are either still essential, or have left without being replaced with nationals of comparable qualifications.
- (7) Overambitious design is an appealing argument, but the study cannot identify it as decisive. All seven projects ran large scale farm field surveys, continuously or intermittently throughout the period of the projects. The results, especially the measurement of yields, have been challenged in all but one case. Nevertheless it is not clear that these surveys invariably put stress on the rest of the M&E agenda. Resources measured by *number* of staff and *size* of budget were adequate to support many other activities in the monitoring category and the results and utilization of these activities was also disappointing. Overload, and the quest for unreachable but staff-intensive targets like yields, probably contributed to deflect the operation from M&E activities with quicker returns.³⁰ But something else explains why neither the big nor the small studies had much impact.
- (8) Finally, location also does not stand out as a determining factor in this review. Of the seven projects, all but one put the monitoring function inside the project, and the other (Nigeria) split it both in and out. However, three of the projects put evaluation outside, including Nigeria, and in none of the three as well as none of the other four has the experience with evaluation been successful. As suggested in point (7), some other conditions were overriding the location factor. One observation on the sample of 172 projects that is germane to this

30. One commentator on the draft of this report believes this to be the unifying factor that doomed all seven programs: the futile and disruptive attempt to measure yield increases. OED feels that conclusion cannot be confirmed on the basis of the information available.

issue is that the use of M&E Units - identifiable parts of the organigram, responsible solely for monitoring and evaluation operations, and popular in the 1960s and 1970s especially in agriculture - has faded almost entirely from the new projects in the sample. AGRME had warned against the establishment of these Units, and the agriculture sector has largely abandoned them in the new designs (they continue in the older project series).

C. Other Sectors

4.31 Ex-post treatment of M&E in PCRs and PARs for the other sectors is weaker than for agriculture. In fact, the information gap is worse, since the old set of SARs called for less M&E and there was correspondingly less to discuss in the ex-post review and less to recall in interview. But, with some exceptions, the PCRs and PARs that do discuss M&E are as uncomplimentary as those in agriculture. The following information is drawn from the sample, supplemented by two case studies of M&E carried out by OED during recent audits.

4.32 For education and PHN, the patterns are similar to each other and no different than agriculture. In terms of PCR and PAR reporting on the performance of M&E and the generation of KPI data, the ratings are again unsatisfactory. For education, for 7 of 11 projects in the old set the commentary - typically sparse - indicated poor performance or at least results less than projected by the SAR. One exception is a real outlier, because in this case very satisfactory performance of M&E and KPI was reported but none of it had been anticipated in the SAR (Philippine Vocational Training Project). A typical case pointing in the opposite direction - plans that did not materialize - is discussed in an OED review of the Education Reform Program in Ghana, where government gave more attention to an ad-hoc monitoring system than to the project M&E (Box 12). For PHN, the record is better. There were four good performers out of ten - including Indonesia Second Nutrition and Medical Health and China Rural Health and Medical Education. In this set the Indonesia project's good performance was anticipated by a strong appearance of M&E in the SAR, while the China SAR, like the Philippine education SAR, is almost silent on all aspects of M&E. As with agriculture, the documents do not analyze the reasons for poor performance, but, where hints are given, they point to low level support from government.

4.33 The status of M&E in three completed urban projects in Ghana and Kenya (outside the Overview sample of ten old urban projects) was investigated during recent OED audit missions. In all cases, the SAR called for M&E but the institutional arrangement was poorly specified and results were minimal. Box 13 discusses the Kenya Secondary Towns Project.

4.34 The data gap for M&E in the old set of projects widens for the rest of the "hard" sectors. In industry, none of the five old SARs had identified KPI or provided for any M&E, and nothing is said about them in the PCRs (except that one of the project PCRs presents a few KPI). For power, four of the seven old projects were for rural electrification, and two of the four had a substantial M&E component written into the SAR for Bank support (Brazil, Thailand). A third referred to a large-scale evaluation study (not by the M&E system) to be financed by another donor (Bangladesh-USAID). In all three cases, the PCRs, PARs and subsequent sector-level reports say that the M&E was either not carried out (Thailand, Bangladesh) or no results had been sent to (or demanded by) the Bank and it was uncertain whether the studies had been completed (Brazil).

4.35 For transport, M&E did not appear anywhere in the ten old SARs, except for casual use of the word monitoring in the section on progress reporting. However, as mentioned above (para 3.16),

KPI have a long tradition in highways and ports (and railways) and in these sub-sectors the SARs as well as the PCRs do report target indicators and results. These were common products of whatever MIS systems were operating, and not the result of any extraordinary M&E activity.

Box 12: GHANA - Monitoring of Education Services

After a decade of deterioration in its education system, Ghana launched a comprehensive education reform program in 1986. Its goals are to replace the previous English-style academic education with a vocational and practical oriented curriculum, reduce the number of years required to complete the primary and secondary cycles, increase access to basic education for the rural masses, improve quality, and accomplish all this in a financially sustainable manner. The program has been heavily supported by the Bank and other donors. Altogether, seven Bank projects have committed US\$232m to assist Ghana's education sector since 1985. As part of that package, the Bank provided funds for a technical assistance (TA) project largely financed by UNDP to support a central Planning, Budgeting, Monitoring and Evaluation (PBM&E) Division established in the Ministry of Education (MOE).

So far, this office has not achieved what was expected of it. It has resulted in some consolidation of planning and budgeting functions. It has produced a number of useful manuals and guidelines on planning, budgeting, statistics and MIS, statistical reports, an enrollment projection model and estimates of unit costs. But it never developed into a high level planning agency, has not undertaken any serious policy analyses or evaluations of on-going operations, and has experienced a serious decline in its level of output since the termination of the technical cooperation agreement in December, 1992.

A draft OED report on Ghana's Education Reform Program speculates on the reasons for these disappointing results. They include start-up delays and personnel problems experienced in the early years of the TA project. Delays were attributable to slow mobilization of the technical assistance by the responsible agency - UNESCO - and the replacement of the team leader a year after his arrival. In the meantime, education decision makers came to rely on the Project Management Unit, running the whole of the Reform program, for policy planning work. Also, PBM&E was never fully and appropriately staffed, the assumption being that TA would fill the gaps. But, initially, the Ministry wanted only short-term consultants, and the expatriate TA team leader was unable to play the leadership role on a visiting basis. Later, when a full-time consultant was appointed, things worked well but only until his contract expired. The OED report concludes that it would be inappropriate to follow with a second TA project before PBM&E is provided with the appropriate professional staff.

But PBM&E has not been the only source of monitoring and evaluation information available to the education decision makers. More important were District Educational Reform Committees, which were served by a special cadre of sub-district based circuit monitoring assistants. These committees reported to a National Implementation Committee headed by the Deputy Minister. This ad hoc system effectively provided the Deputy Minister with relatively accurate national data within a month of its request. Unfortunately, this system did not monitor learning achievements and it never became institutionalized. To correct for the absence of information on learning achievements, the Bank proposed under its Second Education Adjustment Credit, and USAID funded under its primary education project, a "criterion reference test" to determine whether primary school graduates were learning what they were supposed to from the new curriculum. This test, first applied in the 1991/92 academic year, is an important development which promises to provide ongoing monitoring of the learning process at least for the primary level. PBM&E would have been a natural base for this operation, but has had no direct involvement. In fact, PBM&E has concentrated mostly on P and B, leaving a vacuum in M and E to be filled by these other agencies.

4.36 The group of transport projects include two of special interest to the study. The Hungary Transport Project (for both rail and road), approved in 1985, has a well developed KPI treatment in the SAR and good monitoring and KPI performance as reported in the PCR. The PCR calls

attention to the fact that these indicators were intended and used for updating the rate of return analysis in preparing for the next project, the first reference anywhere in these documents to KPI being put to this use (though presumably not the only time they were so used). This KPI experience stands out.

Box 13: Kenya - Secondary Towns Project
Cr.1390 for US\$22.0m/Ln 2319 for US\$7.0m. 1983-1990

This is representative of a broad range of projects where the SAR provided for a monitoring system, supported by special evaluation studies, but the institutional structure was poorly specified and KPI were not identified at appraisal. The project financed site and service development, housing construction loans and community facilities at five rapidly expanding towns. A Project Management Unit was responsible for coordinating the works and monitoring progress. An evaluation unit in the housing ministry was to carry out occasional studies. The SAR called for the establishment of a "Project Monitoring System", including key indicators and discussion of physical works in quarterly progress reports, as well as in-depth and comparative analyses of evaluation topics in selected towns or across all five towns.

The PCR and PAR say that progress reporting was adequate for the first three years, after which both township and central reporting declined, that the "System" was never established, and that no studies were carried out. This was a problem project, there was a struggle among the ministries involved, the PMU had little support from government and lost morale and senior staff after 1986, and US\$14m of the US\$29m package was cancelled. The frustrations at PMU can explain a large part of the lack of follow-through on M&E. The PAR says also that the "System" was badly specified at appraisal and the local consultant recruited to elaborate that plan misread the TOR and worked only on the quarterly reporting format. Supervision complained that it had to prepare its own data base. Government complained that Bank supervision was not helpful. Obviously ineffective management helps explain poor M&E. However, better M&E and KPI would not have been able to control for the political shocks. The Bank was unable to turn the situation around. The PAR also notes that the Bank had expected to use M&E results from two previous urban projects in Kenya to help guide planning and implementation of the new project, but that these results were also disappointing, delayed and unused. ^{iv}

^{iv} It should be noted that the Kenyan Government rates the project as relatively successful, despite underexpenditure in some categories.

4.37 The second project of special interest is the Mexico Fourth Railway Project, which gets good marks in the PCR but is criticized in the PAR on an issue which the PCR did not develop - the decline in rail usage, undermining the economic and financial justification for this and the follow-on projects. The PAR makes the point that the principal Mexican railway authority's MIS system was functioning well, and the secular deterioration in the trend of railroad usage was unmistakable. Yet these data had no substantial impact on the authority's or the Bank's decisions about continuing with the rail expansion program, which proceeded as if usage rates were suffering from only a temporary adjustment. Here is a case of good M&E, and no appropriate management response.

D. Experience of Three Donor Agencies

4.38 The terms of reference for the study called for a review of M&E experience of other developmental aid agencies. OED examined summarily the situation with M&E in two of the donors with the best reputations for promoting ex-post evaluation: the International Fund for Agricultural Development (IFAD) and the United States Agency for International Development (USAID). These remarks are followed by comments on the work of GTZ (Deutsche Gesellschaft fuer Technische

Zusammenarbeit), the German technical assistance agency, in adapting logframe and providing technical assistance to the Bank.

4.39 Before turning to those examples, it is worth reporting that the borrowers have often expressed unhappiness with the level of coordination between the development aid agencies in programming and executing M&E and indicator activities. The main problem has been the multiplicity of donor demands for performance data, each to satisfy its own reporting requirements and sometimes different enough to drive the M&E project staff crazy. At a higher level of M&E assistance, where donors deliberately collaborate in setting up and supporting M&E operations, the relationship is usually productive. The Bank and IFAD have often teamed up successfully on M&E in rural development programs they co-finance (though the illustrations from Africa in Box 11 show that the two agencies can sometimes get way out of step).

(i) IFAD

4.40 IFAD has been busier than the Bank in setting up M&E systems in its rural development projects - all of them intended to benefit exclusively the poorer households. Given such explicit targeting, IFAD considered purposeful monitoring of results essential. It offered its borrowers technical support in establishing M&E units, using short and long term consultants financed by IFAD. (AGRME had offered technical assistance from its own staff, but only through occasional visits - para 2.6). IFAD's Monitoring and Evaluation Division (MED) originally provided this technical support. But in 1988 the responsibility was transferred to IFAD's regional operating divisions. An analysis of the performance of the monitoring and evaluation systems in IFAD supported projects reached a rather depressing conclusion:

"It is clear from this review that there have been considerable problems in the implementation of M&E in IFAD financed projects. M&E designs have often been overambitious, specifying many more activities than the available M&E staff can handle, and requiring complex procedures which untrained and inexperienced staff cannot adequately implement".³¹

4.41 Later, the IFAD report mentions "the shortage of trained and experienced staff" as a dominating problem. The article cited in para 4.17 which presented figures from OED's rural development report also refers to IFAD's experience. It states that:

"this low quality of M&E performance is particularly notable since both the World Bank and IFAD have emphasized their commitment to project M&E and both organizations have committed significant funds to (it).."(page 500).

4.42 IFAD's M&E technical assistance program has declined since the transfer, a casualty of budget shortages and the different priorities of the operational departments. The whole of this story is consistent with that of the Bank.

(ii) USAID

4.43 USAID's evaluation programs are different. USAID normally incorporates a requirement for M&E in the appraisal document for developmental assistance programs, but usually it does not

31. "Analysis of the Performance of the Monitoring and Evaluation Systems/Arrangements in IFAD-Supported Projects." May 1989. IFAD, Loan Implementation Unit. Report #0167. Page 37.

try to build M&E competence in a cooperating host-country agency. When USAID calls for ex-post evaluations, it carries them out through studies by staff of its own Program and Operations Assessment Division (with the help of the staff's American consultants and sometimes with host-country participants). USAID has completed a large number of excellent evaluations. But it can be associated with only a few of the M&E institution building programs familiar to the Bank and IFAD.

4.44 What USAID has done as well is to promote the most extensive use of KPI of any of the donors. For USAID the link between project and program objectives has always been paramount, and the KPI are organized hierarchically to bridge that space. In the 1960s USAID pioneered the use of the Logical Framework, a program and project planning methodology that assigns indicators to each level of expected results - from delivery of the planned inputs through direct outputs to overriding program objectives (para 2.27). USAID spread the use of logframe throughout its overseas missions two decades ago, and other donors (e.g. GTZ) and consultant firms have adopted it as well.

4.45 Since 1991, this planning tool has been accompanied by the introduction of another, grander framework for defining country program objectives, under a "strategic" planning-cum-monitoring approach called PRISM (Program Performance Information for Strategic Management). USAID/Washington first experimented with this approach in its development programs in Africa beginning in 1989, and it is now spreading through the world-wide network of field missions, with technical support from headquarters. PRISM encourages mission management to specify "strategic objectives", defined as "statements of significant, sustainable development results being sought by USAID in a country, which are considered achievable in the medium-term (usually 5-8 years) and for which mission management is willing to be held accountable"³² The strategic objectives are reached through a group of lower order and shorter term "program outcomes", which are themselves to be tied to the monitorable outputs of specific, USAID-financed activities. Actually the link from strategic objectives down to specific activities has not yet been firmed except in the most advanced missions, so that there is still a gap between PRISM and logframe. Nevertheless the intention is to tie it all together as older activities end and new ones are planned. PRISM fits the model of the present American Administration's policy for "reinventing government", for that component which promotes "management by results". Once these systems are complete, they can be used to signal problems that need correction and to identify activities that, although purportedly aligned with the agreed outcomes and objectives, are nevertheless failing to produce expected results.

4.46 But USAID's evaluation activities do not confront one of the biggest problems facing the Bank's (and IFAD's) M&E program. The project logframe and PRISM are viewed primarily as tools for improving USAID's own operations, and ownership levels inside the agency are increasing. The level of government ownership has been minimal, again except in the more advanced missions. It has not even been a systematic USAID objective to secure ownership by its borrowers and grantees. The PRISM approach originated in Washington, and "participation" meant bringing the managers in USAID's country missions on board a new strategic planning and management approach. USAID/Washington has had to sell its overseas missions on the new processes and techniques, and their commitment is essential to make it work. A process of extending these techniques to the countries themselves until recently was not a dominant aim, neither with logframe (although some missions have sought government's involvement, and GTZ has adapted it for just that purpose) nor PRISM. The principal government agencies that were brought into the PRISM exercise were those

32. USAID correspondence to OED dated May 16, 1994.

to which missions had to turn for data. This attitude is changing: current programming directives now encourage greater host country participation in monitoring and evaluation as well as planning processes.

4.47 In some of the strategic plans developed by missions, the hierarchy of monitorable targets reaches such a level of abstraction, and intangible causal linkage, that the effectiveness of PRISM's monitoring system in actually informing management decisions during implementation is questionable. Logframe, working at the project level, has been tested and has earned both praise and criticism. PRISM carries the concept to higher levels of objectives, and runs the risk of generalized objectives, overly-ambitious targets and tenuous causality. When this is the case, using KPI data to "manage by results" is not warranted and may backfire.³³ USAID's evaluation staff is aware of the problem and intent on adjusting PRISM until it leads to a viable and useful instrument for improving management. It intends to use PRISM and evaluation in conjunction, where PRISM flags problems and shortfalls that lead to diagnostic study, corrective action and, only if necessary, amputation.

4.48 To the extent PRISM remains a tool used mostly by USAID, and depends mainly on data generated by mission staffs, this grand adventure with monitorable indicators has limited significance to the Bank. If Next Steps stops with the indicators, and does not pursue with equal energy the development of in-country capacity to generate - and use - a viable data base, then the Bank's experience would resemble USAID's. And neither program may survive.

(iii) GTZ

4.49 USAID had developed logframe for both technical assistance and capital aid projects. GTZ took it up in the early 1980s for use with its own technical assistance to projects implemented by its partners. In addition to the logframe a participatory procedure was developed in order to make sure that the purpose of the project (intended impact) is relevant to the needs of the beneficiaries.

4.50 Consequently the GTZ model, ZOPP, is distinguished from USAID's model by compelling government (or non-government) implementing organizations and beneficiary involvement in successive iterative stages of defining objectives, indicators and assumptions. KfW (Kreditanstalt fuer Wiederaufbau), Germany's capital assistance agency, moved in the same direction, first adopting a version of logframe similar to the philosophy of USAID's, subsequently taking up ZOPP itself (since GTZ had already designed ZOPP to fit capital aid projects too).

4.51 ZOPP is the German acronym for an integrated planning/implementation system labelled "objectives-oriented project planning". Linking planning to implementation and both phases continuously to the intended impact is a primary objective. In that sense it anticipates one of the primary recommendations of PMTF. According to GTZ' training manual for English audiences, ZOPP "forms the basis for project management and is generally used in various planning steps for the purpose of project preparation and project implementation". GTZ made ZOPP mandatory for

33. For example, one of the highest objectives with a monitorable numerical target for a USAID technical assistance project for Burundi's agricultural research program was, in a version that has recently been modified, a percentage increase in the country's coffee and other commercial farm exports. If USAID/Burundi were dedicated to applying "management by results" through PRISM, and had used this indicator (among others), the stagnation of coffee exports would have suggested that it significantly adjust or terminate the technical assistance. Such a mechanistic response, wielding a higher order objective, would make the USAID/Washington staff promoting PRISM - and Bank staff now engaged in the search for KPI under the Next Steps program - shudder.

all of its own management tasks in 1983. In the last two years, ZOPP has spread rapidly also in the Bank, which previously had shown little interest in the logframe methodology. GTZ responded to the interest of individual Bank staff member by providing free technical assistance. The participatory feature of ZOPP, which improves project design and enhances government's commitment, helps explain the Bank's latter-day conversion to logframe methods.

4.52 The main point of entry has been the Sahelian Department of the Bank's Africa Regional Office, which has prepared or already sent to the Board over ten ZOPP-based projects in education, PHN, agriculture and the environment. The model has been picked up by other departments and regions of the Bank. For example, ZOPP has been introduced by GTZ trainers to Brazilian officers responsible for implementing the reformulated PAPP rural development program in the Northeast states (para 4.23). ZOPP is spreading in other donor agencies as well: for example the Commission of the European Union, UNIDO and NORAD each have handbooks or guidelines on using this planning model.

4.53 GTZ puts ZOPP at the center of its project management practices. GTZ also supports the strengthening of management capabilities - with regard to operational planning, monitoring, evaluation and reporting - in the ministries and project executing agencies benefiting from German technical assistance. In that respect GTZ's experience is more relevant to the Bank than USAID's. Nevertheless, GTZ' M&E institution-building activity has not reached anywhere near the size of the operations mounted over the last two decades by IFAD and the Bank. It is important to know whether the monitoring facilities intrinsic to ZOPP have been used adequately by the responsible managers, and to what extent this has resulted in more effective implementation of projects supported by GTZ than by IFAD's and the Bank's. OED did not carry out that comparative analysis for the Overview.

V. RECENT WORK ABOVE THE PROJECT LEVEL: REMARKABLE PROGRESS

A. Sector Indicators

5.1 Recognizing that "good projects depend on good sectors" (as well as visa versa), the PMTF Report calls for the establishment of KPI lists at both the project and sector levels. Most of the early work for the ongoing Next Steps indicator exercise dealt with project KPI. But some good work has also been submitted at the sector level, where a few teams focused their attention from the beginning and where other teams moved subsequently to achieve a better balance.

5.2 Several types of sector level KPI have been identified. The draft KPI submission for the livestock sector, for example, looked for what it refers to as "inventory" and "efficiency" parameters which measure the stock and productivity of the livestock industry as a whole. The industry and mining KPI team took the sector work further by focusing on the exogenous and policy variables that establish the environment in which the project must function. These are the so-called "risk" factors, at the sector level, which, together with the input and "process" indicators measuring project implementation performance, will determine its outcome. To some persons these risk indicators are the key to better management of the portfolio: they define the enabling environment and help identify (1) project proposals that have a low chance of success, as well as (2) the projects under implementation with components that may be undermined by adverse movement in the trend lines represented by the indicator values. The industry and mining KPI submission distinguishes between "referential" and "performance" indicators at both the project and sector levels. The referential indicators are largely exogenous; the performance indicators are subject to control. Taxation policy and barriers to entry are examples of the latter. The team believes the Bank should not attempt investment operations, or continue to support them, if the sector "performance" variables are not under control. It would condition project loans on policy indicators, a practice that is normally restricted to adjustment lending. The analytical treatment of the enabling environment is thorough. In fact, in its first draft of indicators for the mining sector, the team ignored the process-type project implementation variables that many of the other KPI teams were concentrating on. In this case, the team added the project KPI later.

5.3 A much larger exercise has been underway for four years in the urban sector with the Housing Indicators Program, a joint Bank-UNCHS (United Nations Center for Human Settlements) exercise. Here attention also concentrated on the sector level. The main objective has been to identify indicators that establish a framework for discussion of new investments. The extensive KPI lists that have been developed are applicable also to assessing the changes in an enabling environment during project execution. In this case, the housing team in the Transport, Water & Urban Development Department (TWU) distinguishes indicators that serve as *diagnostic* tools for understanding when a sector is working well or badly - and why, and the *policy* indicators that permit monitoring of agreed reforms in policy, institutions and regulatory controls. Property rights and levels of available mortgage finance are examples of policy instruments. As with the industry and mining team, the housing team considers these sector level variables at least as important in predicting and assessing project success as the conventional project indicators. Box 14 describes the application of the Housing Indicator Program's conceptual framework and worldwide averages to South Africa.

Box 14: Housing Indicators Program: South African Housing Sector Performance

The Global Strategy for Shelter calls for a fundamental shift in government's role in housing. Rather than attempting to provide housing directly, governments are called on to play an enabling role. This shift necessarily requires governments to obtain a broader overview of the housing sector as a whole, and to better understand the mechanisms governing housing sector performance. The need for better sector data and for operational tools to measure performance led, in 1990, to the creation of the Housing Indicators Program. A survey of 52 cities in as many countries resulted in the identification of 10 key indicators found to be the most instructive - for assessing housing sector conditions and comparing them across countries - and suitable for regular collection. In 1993 the Bank's Urban Development Division used this analytical framework to describe housing conditions in South Africa.

The report on the study begins: "housing in South Africa is as good as the world has to offer, and as bad."³⁴ It reviews in turn each of the 10 indicators, revealing the striking contrasts between the conditions of the white and black communities. For the House-Price-to-Income-Ratio, for example, it cites the unusually low ratio of 1.7 for the area surveyed (the Pretoria-Witwatersrand-Vereeniging area, PWV). This compares with the world-wide mean of 5.0, and is judged to be the result of severely depressed demand compared to other similar countries:

"The major reason for suppressed demand, however, stems from the effects of apartheid. Because of the spatially dispersed patterns of settlements brought about under apartheid, pressures on land values near city centers and places of employment have been much lower than would otherwise have been the case. For white South Africans, this has led to comparatively lower housing prices and more housing consumption, but for non-white South Africans, it has resulted in what are in effect a wide range of taxes on housing, which have profoundly depressed the demand for housing. The most obvious of these taxes is that associated with the excessive transportation costs and commuting times borne by black South Africans" (page 8).

The other indicators show similar results. Housing investment in the PWV area in 1990 was estimated to be the lowest relative to gross city product of any of the 52 cities surveyed. The median floor area per person shows extreme variation among different population groups: the whites typically enjoying 33m² per person, blacks in the informal sector about 5m². The report concludes that the housing sector is neither serving the interests of the black population nor the economy as a whole. It recommends changes in policies "to achieve the full potential of a well-functioning, fair, and self-sustaining housing sector" (page 19). This exercise illustrates the potential of the Housing Indicator Program - providing a proven analytical tool for describing and explaining country situations in relation to world-wide norms.

³⁴ "South African Housing Sector Performance in International Perspective." Mayo, Stephen K. Urban Development Division. 1993. Page 1.

5.4 The most impressive indicator report prepared for the Next Steps is in the super-sector of poverty programs. The draft report - *Performance Indicators in Poverty Reduction Operations*³⁴ - discusses the purposes and proposals of the PMTF Report, the nomenclature and characteristics of indicators, and the extensive, annotated indicator lists developed by the authors. It stands as a major contribution to the literature on evaluation and indicators.

34. Carvalho and White, op.cit.

5.5 Throughout the 1980s the little work that was done on KPI was uncoordinated, with no sharing of ideas and conceptual structures across sectors. The Next Steps exercise has focused the Bank's attention on this management tool and brought the initiatives together.

B. Participatory Evaluation

5.6 The increased interest in the Bank in participatory activities, spearheaded by the Learning Group on Participation formed in December 1990, interacts with the sector and project M&E activity. In particular, participation is encouraged in the M&E process itself. A distinction has to be made between "beneficiary participation" and "participatory evaluation". Some staff in the Bank use the two phrases interchangeably, although a convention is emerging to use the first phrase when seeking beneficiary opinions about project design and progress (captured in the title of the book *Listen to the People* by L. Salmen³⁵), and reserve the other phrase for the actual involvement of beneficiaries in managing the M&E process. The latter includes not only designing the M&E system and agreeing on the indicators, but also monitoring the implementation and impact of projects and helping find the lessons and put them to work. Both phrases are defined and discussed in the paper *Monitoring and Evaluating Popular Participation in World Bank-Assisted Projects* submitted in 1992 to the Bank's participation workshop.³⁶

5.7 Whereas "beneficiary participation" is taken for granted now in all new projects in the social program sectors, "participatory evaluation" (PE) was built into only seven of the SARs included in OED's scan of 83 new projects. This is not yet an established methodology in the majority of the social program sectors, including agriculture. It will be important to monitor closely the effectiveness of PE in strengthening M&E in those few new projects highlighted in Chapter III which have a high PE content - for example the involvement of Egyptian Bedouin communities in monitoring progress in managing their own cropping and livestock resources (Box 6). OED visited the two water supply and sanitation projects that feature PE as a component of structured learning (para 5.12), and found that the participatory elements of project design exaggerate the level of involvement of beneficiaries in monitoring in practice. But the programs are too new to draw any conclusions yet about the ultimate effectiveness of PE in these two applications. Nevertheless, there is a body of experience emerging about successful management of PE in certain sectors, including WSS. One expression of this is the 1993 Technical Paper *Participatory Evaluation, Tools for Managing Change in Water and Sanitation*³⁷. This is not a theoretical treatise: the author argues that the volume "represents the lessons learned in fifteen years of work in participatory development". With respect to PE it is a path-breaker for the Bank in terms of popularizing the tools and potential for community involvement. It does not address directly the demand side of PE: how to set up incentives to persuade the communities to maintain their interest after the change-agents have left. But it suggests that attractive tools provide their own incentives. Some of the techniques recommended in the paper for securing beneficiary collaboration in M&E are described in Box 15.

35. "Listen to the People, Participant-Observer Evaluation of Development Projects." Salmen, Lawrence F. 1987. Oxford University Press.

36. Uphoff, Norman. In "Participatory Development and the World Bank, Potential Directions for Change". Ed: Bhatnagar, Bhuvan and Williams, Aubrey, C. World Bank Discussion Papers 183. 1992.

37. Narayan, Deepa. Technical Paper 207. 1993.

Box 15: Participatory Evaluation in Water and Sanitation

World Bank Technical Paper Number 207 - Participatory Evaluation, Tools for Managing Change in Water and Sanitation^{2/} - is a fascinating, illustrated study of "ideas about participatory processes and indicators that can be used to involve the community members and others in program evaluation". The author was then manager of the PROWESS program of the UNDP-World Bank Water and Sanitation Program. Her objective was simple:

"Over the years, there have been dramatic changes in the way development projects are planned. Participatory planning is now widely recognized as more likely to lead to designs and strategies that work in the particular setting for which they were intended. However participatory data collection for monitoring and evaluation is not yet an integral part of the development process. When it comes to evaluating projects, there is still great reluctance to move from classic "objective" methodologies that maintain a distance from the people and activities being evaluated. There is surely a place for the classical approach. But when the goal is to enhance local capacity, it is of limited value to have an evaluation process directed by outsiders and which generates reports which may not be disseminated for months or years. My hope is that this document will help rectify this situation by moving participatory evaluation into the mainstream of development" (page ix).

With numerous diagrams and photographs, she then discusses and illustrates practical methods for collecting field data for each of 33 indicators of progress in water and sanitation programs, indicators divided into three over-arching categories of sustainability, effective use, and replicability. Matches and match boxes, drawings on the dirt, cut-out human bodies with movable limbs, and other simple devices are part of the tool kit. Children are asked to draw pictures of their village, and draw as well the problems that need to be tackled. The report warns against overly sophisticated analysis: the difference between mean family sizes of 6.1 and 6.7 is said not to matter much to a community intent on promoting the use of household latrines. The report pulls women into the middle of the evaluation activities.

This study straddles the borderline between beneficiary participation and participatory evaluation (see para 5.6 of the main text). Many of the tools are presented as devices to get the villagers attention and good answers. But the tools are user friendly, and can be put to work in subsequent rounds by the participants themselves. That is the author's real agenda. The report is recommended for anyone interested in strengthening M&E in community-oriented projects. It is recommended also for anyone who visits villages without appreciating the capacity of the villagers to express what is on their minds.

^{2/} Narayan, Deepa. 1993

5.8 Participatory evaluation must be distinguished also from the practice of involving government and project staff in identifying objectives and monitorable indicators. This is already commonplace in preparation and appraisal work. It is an essential part of the architecture of the PMTF management system. Two particularly good examples of Bank efforts to involve governments in developing the KPI lists are found in the China and Brazil projects included in the sample of 1993 SARs and described in Boxes 4 and 5. In the China Sixth Railway Project, the task manager wanted monitorable indicators for each of the project's six targeted objectives. But he let the six groups of Chinese involved in project design develop their own (extensive) lists of indicators and an M&E plan to collect the data. Similarly, in the two Brazil Northeast Basic Education Projects, the task manager wanted a short set of three "trigger" indicators for each state - to permit release of annual Bank funding for the school construction component. That prompted a three year preparation exercise with full involvement of state officials and technicians to get agreement on project design, monitorable indicators, and a plan to put the indicators into play. The Brazilian projects offer a

striking contrast between KPI prepared with government involvement and KPI prepared with no involvement. One annex in each of the two SARs describes the three "trigger" indicators, which have the state governors' approval. Another annex in each SAR contains many tables of other input, process and output KPI. These annexes were prepared in the Bank without state participation, the state governments claim little or no knowledge of them, but the Bank is currently financing a consultant team to work with the state governments to operationalize the lists (see also Box 5). For these KPI, the Bank has to work after the projects have been negotiated to secure government ownership.

C. Recasting M&E in the Bank: The Learning Culture

5.9 Modern books about corporate behavior do not talk about M&E, they talk about learning organizations. Commercial success in an increasingly competitive world will depend upon steady improvement in product design, quality and acceptability. These, in turn, call for a corporate culture that rewards continuous monitoring and evaluation of practices all along the production line and rapid feedback of lessons. W. Edwards Deming, the inspiration of the quality revolution, preferred to be remembered for "continuous process improvement" rather than "quality control". Employees at each level are expected to "participate", though that word does not adequately capture the spirit of a learning culture which becomes second nature to every individual on the team and where opinions and suggestions for improvement are not so much encouraged as expected. Learning organizations do not have to establish an M&E system. It is part of the routine.

5.10 The movement is gathering support, first in Japan under the tutelage of Deming and now in the United States and Western Europe. In 1980, D.C. Korton wrote one of the first articles on the new science relating it to development work, arguing that the development process itself should be viewed as a learning experience for all the participants³⁸. He compared the "blueprint" approach ingrained in Bank work with the "learning process", and encouraged practitioners to "embrace error". Some U.S. corporations have embraced the new orientation (Peter M. Senge's 1990 book *The Fifth Discipline* gives examples). The Chief Information Officer from McKinsey and Co., one of the leading proponents of the new learning paradigm, presented "the vision of the knowledge organization" to a Bank audience in June 1992. The U.S. Government is moving fast in this direction. The Administration's focus on "reinventing government" incorporates the features of the learning culture among other innovative behavioral changes.

5.11 In the Bank as well one sees early signs of a shift in the same direction. The most important of these initiatives is the Southern Africa Department's (AF6) decision in 1992 to remodel its operating system along the lines of the quality seeking, learning model. The adjustment is Department-wide, and it effects as well the way business is handled with the Borrowers. It sets AF6 apart from other departments in the Bank, although the Regional Vice President endorses the philosophy and is encouraging his other departments to consider the innovations. Another initiative in the Bank is the coalescing of an informal and expanding group of staff in a "quality network". Meeting once a month, with lively EM networking in-between monthly sessions, these persons seek to promote improved quality throughout Bank business and in particular of the portfolio. The learning theme permeates this network: the concept being to build on each others' experience.

5.12 Independently of those activities, TWU is promoting throughout its linked sector operation divisions the concept of "structured learning". That phrase has been in common use in the education

38. "Community Organization and Rural Development: A Learning Process Approach." Korten, David, C. In *Public Administration Review*. September/October 1980.

profession for decades. The quotation marks signal its special usage in TWU. The idea crystallized in 1992 within its Water and Sanitation Division, which wanted to see a shift in project design to a more experimental mode in the rural and urban water supply projects dealing with lower income households.

5.13 Two projects - the PROSANEAR project in Brazil, approved in another form in 1988 and reformatted in 1992, and the Water Supply and Sanitation for Low Income Communities Project in Indonesia, approved in 1993, are already under implementation. The distinctive feature of the two projects is that the communities (Brazil) or provincial authorities (Indonesia) select from a menu of alternative strategies for building, maintaining and monitoring the low-income urban (Brazil) and rural (Indonesia) WSS systems. Assuming the communities do not all choose the same option, the chance to compare costs and results is a valuable by-product and part of the learning experience. But the projects do not enforce diversity, and the major learning feature of the operation is the improvements expected as the participants - project staff as well as communities - deal with problems and realize the benefits of self-selected strategies. Participation is inherent in SL. So is monitoring. The hope is that these become the routine, as in the corporate paradigm, and will "replace the prevailing normative model of project design with a flexible, adoptive learning process"³⁹. Thus the two SARs are steeped in the language of M&E and operational, demand driven, indicators. These water supply and sanitation projects are described further in Box 16.

5.14 As mentioned above, participatory evaluation is a normal feature of structured learning. It is the lowest layer of the learning culture as that concept is applied to a Bank project. Project management units at scheme and state levels would be interacting with project beneficiaries and among themselves in a continuing, iterative process of observations and lessons moving up and down the ladder. The Bank in theory sits astride the process, learning from the experience of all groups.

5.15 When such a process is an integral part of the project, M&E is taken care of. One does not have to set it up as a separate function. In the ideal situation, the demand as well as the supply side of M&E are taken care of. The participants look for and make use of the lessons of others, and readily share their own. In a sense, M&E is a contrived form of the learning process that is gradually absorbed by the culture as it takes over the routine. SL is moving in that direction. So is AF6. One of the M&E experts interviewed by OED during the study said that the goal of the process ought to be just that, to internalize the M&E functions so they become invisible. He felt that where you found an operating M&E system with that name, it was either new or should be dismissed as a "basket case". The importance of this process of not only mainstreaming but internalizing the M&E, or learning, culture is caught in the title to this report. The Overview was deliberately aimed at M&E "in the Bank", rather than "in Bank projects", because the cultural shift is more important and will guarantee that the second comes along as well.

5.16 An interesting innovation in Bank appraisal work is worth reporting at this point. In 1993 a task manager drafted an appraisal report package for the "Iran Rangelands and Livestock Development Project". He adopted the appraisal format proposed, but not yet approved, under the BIAS program. He was the first to do so. The proposed format would dispense with the SAR and replace it with an expanded President's Memorandum and a Project Document (PD). This format has allowed him to shift priority to project management rather than Bank readers. The PD becomes a tool of implementation. It includes detailed implementation schedules and terms of reference for interim reviews. It also makes heavy use of monitoring indicators and local M&E consultants.

39. "Structural Learning: A Discussion Note". Briscoe, John. TWUWS. January 9, 1993, draft.

Project processing was subsequently held up, to convert the documentation to the traditional format. The reversal is explained by political factors unique to this Bank/Borrower relationship. Nevertheless, the M&E content was preserved. Also, interest in the proposed changes in format are gaining momentum and OPR has drafted instructions for the Bank to use the PD approach in other, pilot, operations in 1995.

5.17 There are two characteristics of the learning paradigm which will slow its progress in reaching the objective: an easy and iterative sharing of experience spreading within and between the Bank, Borrowers and beneficiaries. First, participatory evaluation will have to be guided. Villagers are accustomed to collective decision making, but not in projects imported from outside the community. A tendency for communities participating at the project *design* stage to present unattainable wish lists has to be anticipated. Similarly, community participation in *monitoring* the use of externally provided inputs is unlikely to lead readily to self-examination and self control.

5.18 This was evident in OED's field trip to inspect the "condominial" sewerage projects in the northeastern cities of Brazil, some financed under the PROSANEAR project (para 5.13). In this part of Brazil, the authorities feel that the lower-income communities are not yet prepared to control the system themselves, a function which *has* been taken over in cities in the more developed part of the country. Some form of external authority and intervention must be maintained. That is not incompatible with the SL concepts. These easily allow for a division of labor in managing the assets of the water supply system, where the water company would forever be responsible for maintaining the trunk lines, and the community - or at first perhaps an NGO representing it - would maintain the communal feeder lines with the company on call to assist as needed. Over time, SL predicts that the frequency of calls would diminish.

5.19 The second characteristic is that the learning culture, and the internalizing of M&E, are alien also to most Bank and government staff, and the full transition to the new culture is still a distant vision. The concept of "embracing error" drew criticism during review of the draft of this report, from persons who do not deny its attractions but are skeptical that the majority of either Bank staff or government officials will find the new discipline comfortable any time in the near future. They believe that incentives will have to be applied to maintain M&E momentum, and that it is the absence of special incentives - to offset a natural reluctance for people to "embrace" let alone publicize their errors - that explains the disappointing record of M&E in the past. This implies that M&E will continue to have to be exposed (and imposed) as a separate project component - rather than an integral part of the implementation process - for a long time to come.

5.20 A good idea of the distance to travel is reflected by OED's own decisions on handling, for purposes of this study, the spate of environmental projects and components that have appeared in the last few years. When OED first set about identifying the sectors to include and exclude from the study, no separate identity was given to the environmental group of projects, and environmental components in larger projects were treated as exceptional. Invariably the projects and components included M&E - measuring the physical emissions of toxic substances, the numbers of wildlife protected, and other qualities of the human habitat and ecology. OED took it for granted that environmental activities necessarily carried their own M&E culture. Thus, M&E in the environmental component of a larger project was expected to be present in all cases. As this assumption proved to be correct, OED did not bother for the study to rate the environmental components or projects for M&E content but directed its attention to the other components and projects to measure the M&E load.

Box 16: Structured Learning in Water and Sanitation

OED visited selected sites of the two Brazilian and Indonesian water supply/sanitation (WSS) projects that have come under the influence of the structured learning (SL) concept. In Brazil, the Caixa Economica Federal, a public bank, is financing in several cities the expansion of the urban water supply system in low-income neighborhoods. The preferred model for these areas is what the Brazilians call the "condominial" system. It runs common pipes through back or front yards collecting sewage from the whole block before conveying it to trunk lines under the main streets. Cost savings compared with the "conventional" sewerage system derive from avoiding individual household connections direct to the trunk line. (The condominium image refers to an upright apartment building, where dwellers take it for granted their "neighbors" on the upper floors send waste through their walls.) Caixa helps finance the expansion, normally carried out by the state water corporations. Although the condominial program is a decade old, Caixa still regards it as experimental. The piping technologies have been modified, the choice between back yards, front yards, and sidewalks are left to the communities and here too lessons are accumulating throughout the country. The most important learning experience for Caixa, and particularly for the corporations, is in dealing with the communities. Sociologists and NGO's have entered the projects. In the condominial zones in cities in the more developed, better educated southern parts of Brazil, the communities are involved in monitoring household use and system performance, and manage their own repairs. In the northeast, where OED paid its visits, the corporations maintain control. They argue the communities are not prepared to manage collectively the assets, or at least that the risks of mismanagement and system degradation are higher than in the south. However, the pattern that is expected to characterize the majority of the programs is one of a reliable symbiosis of formal and non-formal arrangements, with the companies and the communities sharing responsibilities. The most notable feature of this whole enterprise is the commitment at Caixa to adjust designs and works to the errors and lessons of experience. This is the working model of structured learning that first caught the attention of the Bank.

In Indonesia the program for rural community WSS is younger than in Brazil, project start-up was delayed into 1994, and the initial works by the provincial water authority in Central Java ignored the prescription for community involvement in design decisions. Nevertheless CARE and some other NGOs have a long experience in similar projects, and the Bank is comfortable that the principles of SL will be respected. The model differs from Brazil in that the basic technology is a cheap water supply system, usually tapping nearby upland springs, with standpipes in the yards. Project officers promote the concomitant construction of simple latrines, and of running small pipes from the standpipe to the outhouse for washing hands. There are none of the shared pipeline layout issues associated with the condominial scheme. Nevertheless, here too the Bank intends to keep the program flexible, and adjust procedures, particularly for managing installed systems, as lessons appear. CARE is collaborating with the Bank in several provinces, and brings to the project its own broad experience in other villages. The SAR mentions CARE's use of participatory evaluation, to let the villagers help guide the program. But OED found that the level of community participation in monitoring and evaluation functions was low. CARE argues that community involvement in planning the system layout helps guarantee community involvement in maintenance. But it says that its hopes of building viable, formal groups based on the water systems, and of getting members to contribute to a repair fund before minor breakdowns and emergencies occur, have both been disappointed. The villagers do pay, but not before they have to.

5.21 From the perspective of the learning culture, that distinction makes no sense. The OED study team had to learn that the distinction was not only invalid, but that the attitude that led to it was part of the problem. In principal, there is no more reason to expect environmental projects to carry M&E along with them than for agricultural or educational projects to do so. The individuals in the Bank responsible for the spontaneous emergence of M&E in the new projects take it for granted that the results of their agricultural and educational investments have to be monitored and evaluated as routinely as their environmental colleagues do.

5.22 Most Bank staff do not share those habits. For them, M&E will also have to be encouraged and guided. Volunteer flowering of more and more of these enlightened projects will spread gradually, but it must be nourished.

5.23 It should be noted that the environment staff have to respect Operational Directive 4.01 on environmental assessment, which adds weight to the other instructions on monitoring. That explicit, sector specific pressure helps explain widespread compliance. A discussion of the environmental OD and the Bank's response is given in Box 17.

Box 17: Monitoring Project Environmental Impacts

Since late 1989, all new Bank investment projects have been screened for their potential environmental consequences, and those judged to have potentially "significant adverse impacts that may be sensitive, irreversible, and diverse" (Category A) have been subject to full environmental assessment (EA).^{iv} Projects expected to have less severe impacts (Category B) are submitted to environmental analysis, while those not expected to have environmental consequences (Category C) are not. From the time the Bank's original Operational Directive on environmental assessment was issued through the end of fiscal 1994, close to 10 percent of all operations approved by the Board of Directors -- some 80 projects -- have required full EAs, while another 40-45 percent have required at least some form of environmental analysis.

In addition to the identification and analysis of likely project impacts, full environmental assessments normally include plans both to mitigate and to monitor these impacts. According to Bank EA guidelines, monitoring plans are expected to clearly specify the types of environmental monitoring to be undertaken, who would do it, how much this would cost, and what other inputs (e.g., training) would be necessary. Both monitoring and mitigation plans, moreover, are expected to be fully integrated into overall project design. While formal environmental monitoring plans are not required for Category B projects, many in fact have included monitoring together with proposed mitigation measures.

A typical example of a monitoring plan in a Category A project is that proposed for the Calub Gas Development Project in Ethiopia. This plan consists of both a natural resources and a socio-economic component. The natural resources component will monitor changes in the local biophysical environment including vegetation composition, devegetation, water availability and quality, livestock movements and numbers, and roadside erosion and gulying. The socio-economic component will monitor changes in the human environment including population density and distribution, availability of grains and other goods in local markets, livestock-grain terms of trade, economic activities, health and education, traffic along a road to be rehabilitated under the project, fuel use patterns and prices, and local attitudes toward the project. Monitoring would occur on a continuous basis throughout the life of the project, a specific unit would be established to coordinate monitoring activity, and a five-year budget of US\$ 850,000 is included in the project with IDA funding.

A review of Bank EA experience over the past two fiscal years has found both that the quality of project-specific environmental monitoring plans is generally good and that it has improved substantially when compared with that of similar documents prepared in connection with the first generation of Bank "EA projects" which were surveyed in 1992. The experience with such plans during project execution remains to be seen, however, since most of the active Bank operations that have been subject to full environmental assessment are still in the early stages of implementation. This notwithstanding, the systematic monitoring of key environmental parameters, including air and water quality, is increasingly recognized as an essential tool for assessing actual project impacts.

^{iv} O.D. 4.00, Annex A of October 1989 was later revised and reissued as O.D. 4.01 in October 1991. It is presently in the process of being reformatted and will reappear as OP/BP/GP 4.01 following the new format for Bank internal policy directives. For additional information on project screening and classification for EA purposes, see Environmental Assessment Sourcebook Update No. 2, April 1993.

VI. FINDINGS AND RECOMMENDATIONS

A. Findings

6.1 Inadequate Compliance. The clearest finding is the low priority accorded to M&E during appraisal and implementation despite 20 years of intermittent management exhortation. This gap is especially pronounced for the social programs, where continuing review of performance and feedback into operational plans are essential. Despite agreement at an intellectual level that the monitoring and evaluation functions are important, Bank staff, governments and implementing agencies are not carrying over these commitments into project design, execution and supervision. The Bank issued the two Operational Directives in 1989 with mandatory guidelines on M&E and KPI. The action has not made a significant difference yet on supervision, although the percentage of new projects with substantial M&E content in the appraisal design is rising. "Compliance" refers here to serious and professional efforts to develop M&E systems and indicators. The practice of inserting short promissory notes on M&E and KPI in the SARs is still observed, though some SARs do not even offer this. Most task managers intend that the project include an M&E system, or at least that the MIS system provide the key indicators. But in some cases these intentions are not discussed adequately or at all in the SAR, and in other cases - the majority - the SAR design is either unsophisticated or undermined by inattention to institutional and staffing support. During supervision, these good intentions tend to lapse, and by the time of the PCR they are usually buried.

6.2 Low Level of Ownership. Inadequate commitment is the root of non-compliance, and applies to the Bank as well as to government and its implementing agencies. The prevailing pattern is for the country to look upon M&E and KPI as Bank instruments, and Bank staff to undervalue them because it is unsure how to design and use them and feels little pressure to invest time in doing so. The M&E disciplines are still too fragile to warrant any assumption that their benefits are self-evident and compelling enough to maintain project management's and government's commitment. The belief that "they will do it because it is their interest" has no historical basis. And in both the country and the Bank, there is skepticism over whether M&E and KPI have proven or can prove their value in practice. This helps explain why many governments are reluctant to agree to borrow to finance technical assistance and staff positions for M&E operations. Thus both the demand and supply side of ownership must be promoted. Managers must want to see and apply the products of M&E, and practitioners must get better at producing them. Reflecting the improving trend mentioned in the last paragraph, ownership of M&E is greater in the new set of projects. There are even more cases of the Bank working with the government to develop lists of monitorable KPI. But the numbers still include less than half the projects that should be given such treatment.

6.3 Broad agreement within the Bank. OED found early in the interviews that - apart from the few veteran M&E practitioners - low compliance and ownership was considered the institutional norm. Some sub-sectors are outside that profile, but they really are exceptional. The dominant opinion is not that the Bank is disinterested in M&E, but that it does not know how to carry it out effectively, or to put KPI to good use, and that the countries in any case would not or could not commit qualified staff and adequate resources to these operations. This is a remarkable contradiction, given the equally widespread opinion that M&E and its indicators were essential components of good project management. The early rounds of interviews - spread over all sectors - came to rapid closure on this finding.

6.4 Some Encouraging Exceptions Standing apart are a small group of sophisticated M&E programs included in the sample set of new projects, or in projects reviewed independently by OED

after consultation with staff. They are more impressive than most of the SARs of the old set rated for high ("substantial") M&E content, because the newer M&E activities appear to be better integrated in the implementation process, the search for errors, lessons and feedback is explicit, the emphasis is clearly on quality, participation is expected, and the resources usually match the expectations. These features were not dominant in the "substantial" M&E designs of the projects of the early 1980s. But the more remarkable characteristic of this group of exceptions is that they are scattered throughout the operating divisions, are concentrated in the social programs but break out also in the infrastructure sectors, and exhibit a wide variety of design. They emerge from isolated initiatives by individual project staff personally committed to equipping the project and Bank supervision with a capacity for self-assessment. This is the "learning culture" at work. Other staff within the same divisions are not keeping up with these pioneers. Nevertheless, the chiefs, with a push from many of the project advisors, keep the individual initiatives moving. The movement - if this wild flowering of good M&E can be considered a trend - is not directly related to the Bank's Next Steps effort on indicators, though the sector KPI teams are picking up the names of these best M&E practitioners and incorporating their ideas.

6.5 Progress with Sector Indicators. Chapter V describes the remarkably good work which is underway in several sectors in developing theoretical frameworks for indicators that can track inventories, efficiencies, exogenous risks, endogenous policy variables (that also put the project at risk), and diagnostic sector variables that help steer project planning and implementation. Some of this work predates PMTF, but it has been given a big boost by the Task Force's report and Next Steps.

6.6 Weak institutional support. The wild flowering of the learning culture in project M&E, and the conceptual advances at the sector indicator level, are the more dramatic when set against the vacuum in the Bank for supporting these initiatives with professional advice. The Bank as an organization has talked grandly about M&E during the last 20 years, and urged its Borrowers to take it on. But the Bank has not backed up those declarations with a capable institutional support system. For other movements in the Bank with strong policy support - participation, environmental concerns, women in development - the Bank has established institutional processes to push the action along and give it intellectual power. There is no corresponding support for M&E. OED does not do this at the project level, though it expects ECDP will encourage governments to strengthen project M&E. OPR is organizing the institutional response to the Task Force recommendations, and by promoting an exchange of ideas has helped advance the intellectual content of indicator work. But OPR at present has no plans to set up a permanent clearing house for future indicator work: for institutionalizing these advances under Next Steps. Even the EDI training programs on M&E have been terminated.

6.7 This profile gives the impression that the Bank considers M&E easy: that any task manager after reflection ought to be able to produce appropriate KPI and an M&E format to deliver them. That impression must be corrected. KPI, especially when derived at the project planning stage, may be self-evident. But the M&E institutional context for delivering them, the appreciation of which of the indicators are practical ("do-able"), and the engineering of rapid as well as complex surveys to generate valid impact data are anything but familiar. Task managers and their division chiefs need training and professional guidance. As of now, neither the regional offices nor the central vice presidencies provide it.

6.8 Lack of Capacity Building Assistance in the Countries. The same condition prevails in the borrower institutions, at central and project levels. The Bank calls for consultant support for the projects to reinforce the M&E programs, but this is no substitute for creating domestic capacity. The consultant option is an *alternative* to promoting the learning culture. It will be more difficult to correct the M&E deficit at the country level than it will be at the Bank. While a large majority of

Bank task managers believe that, when properly designed and implemented, M&E will pay-off in improved performance, ministry and project staff in the countries are mostly untutored in the subject, indifferent, or nervous . That is why the ECDP initiative is timely, unique, and important.

6.9 Monitoring versus Evaluation. A difference is observed between progress in these two activities at the project level. Where monitoring measures progress against implementation targets - the so-called input and process indicators - and achievement of obvious and easily identified first-order outputs, the Bank's experience with data collection is reasonably good across all sectors. This progress may not have been written up in PCRs and PARs, but it was put in the quarterly reports and supervision reports and served the monitoring function intended. MIS permeates this scene. Whether an overt component of appraisal, or a quiet component of normal management, the information flows have been adequate to support monitoring devices created to track project progress.

6.10 The weakness appears in the other end of the spectrum of monitoring functions - where information on higher-order outputs and movement toward ultimate objectives is required. This monitoring assignment feeds into the evaluation function, and staff and outside experts are expected to analyze the information provided by the monitoring system and identify constraints on implementation and effects on project beneficiaries. This part of the spectrum of M&E activities is one of the two areas where the Bank's programs have fallen short: in monitoring the higher order output indicators and analyzing project impact. The distinction between areas in the range of M&E activities is important, because it focuses Bank attention on the areas that need help. Much of the routine monitoring activity in Bank projects is on course.

6.11 A draft of this report referred to a "continuum" of M&E activities, ranging from routine financial and physical data collection under an MIS system - on the left end of the continuum, so to speak - through monitoring of secondary and higher levels of output to the analysis and evaluation of final impact - on the other end. Many M&E practitioners commented that the implicit right-handed emphasis pushes M&E Units into studies they are usually unable to do well, and which are of low interest to project management. Also the continuum concept does not admit a discontinuity between normal management monitoring functions, and the evaluation of higher order outputs. The problem is aggravated, the critics say, when the continuum is turned up on a vertical axis. Then the evaluation functions are stacked on top of the monitoring functions, and the sense of an inherent priority on E is reinforced. This is thought to be one of the weaknesses in the logframe structure, which is also vertical, and which puts final impact on "top" of the list of priorities for an M&E operation. It plays to the M&E Unit's weakest hand. For purposes of this report, the word continuum was dropped. The M and E functions are considered distinct but both important, and overemphasis on either will disable the overall effort.

6.12 Enabling and other risk indicators. The other area where the Bank's M&E program has fallen short is in tracking policy and exogenous variables that have a decisive influence on project performance. As mentioned in Chapter V, some staff consider these "risks" to be the missing dimension in project analysis and monitoring. The deficit spreads across all sectors. These indicators distinguish themselves from other project KPI because they call for data from sources outside the project and accessible to the Bank by other routes. The Bank must be concerned with building an institutional capacity to deliver this information. It usually requires dealing with units other than the project executing agencies, and with global and national data sources.

6.13 Variations Between Sectors. The points made in the last two paragraphs can be used to distinguish relative performance across the sectors. The general impression of poorly performing

M&E does not apply everywhere. The distinctions between process and enabling indicators, and between first and higher order outputs and impact, helps separate sectors by level of M&E performance. The Next Steps indicator exercise has brought these differences into sharper relief.

6.14 For the extractive industries, processing industries, energy, utility and other infrastructure sectors, the monitoring programs for processes and outputs are adequate. Further, the standard financial indicators of profitability provide a first approximation of welfare effects. The M&E deficits in these sectors are, first, in developing KPI to track the enabling environment (which is being addressed under Next Steps), and, second, in expanding evaluative research on welfare and environmental impact. Independent evaluative research is preferred to project evaluation routines for assessing the wider effects of these categories of investments. In a ports project, for example, quicker turnaround time can be measured by the project authority, but impact on commercial activity in the hinterland of the port is not an indicator available to that authority. These effects should in any case be investigated only for selected projects. For the oil and gas sector, the PMTF is not asking that the project authorities demonstrate the link to poverty relief and other Bank objectives. What this sector must demonstrate is efficient production and distribution of energy. *That* the existing indicators can handle (even though they do not appear in the SARs). For the social program sectors, however, an M&E deficit is recognized even at the output levels. PMTF insists that the Bank has to get better at assessing morbidity, agricultural production, quality of education, etc. The distinctions between output, impact and enabling indicators is less important here: improvements are needed in all of them.

6.15 This is an important perspective from the Overview. The study intended to look at M&E in all sectors, with a few exceptions. It is clear now that the "issue" of M&E is much more concentrated. To deal with it, a few sectors, and a few sub-sectors in other sectors - almost all of them falling into the domain of the social programs - warrant disproportionate attention. That is no surprise. That is where M&E activity started in the Bank in the 1970s. The "risk" factor introduces another round of M&E/KPI activity for all sectors, but, as mentioned above, it requires separate corrective measures and other institutions.

6.16 **Changes in Character and Clients of the Portfolio.** This said, many infrastructure projects are now viewed as social and environmental interventions and, therefore, require M&E upgrading beyond the traditional indicators. In the utilities, programs are concentrating investments towards poorer communities where subsidies are more common and the market test is no longer adequate. For example, as free sanitation facilities are extended into the chaotic grids of the favelas, water companies are obliged to abandon the market test and verify by on-site inspection that the new facilities are being properly used and maintained. As rural tracks are upgraded to roads, but traffic is as yet too light to warrant the use in economic analysis of the standard test of savings on vehicle operating costs, the road authorities have to assure that the criteria for selecting roads based on expected economic performance are validated in practice.⁴⁰ The shift from new investments to O&M (operations and maintenance) is another trend pushing M&E into more difficult lines of inquiry.

40. Another example from the infrastructure sectors is moving in the opposite direction. In rural electrification, the Bank and other donors have been retreating from the high level of construction activity in the late 1970s precisely because evaluation studies have been showing that the alleged production and commercial impact of RE has been overstated. The only significant benefit of most of these programs is household lighting (the electrification of small scale irrigation pumps in India is an important exception). That is good for poverty eradication, but it is not as good an economic outcome as once believed.

6.17 Similar changes are underway in the social programs. In education, for example, the emphasis has turned from expansion of teaching facilities to improvement in education quality - textbooks, training, teaching, administration, etc. These activities are more difficult to monitor than construction of classrooms.⁴¹ In agriculture, with the reorientation of the portfolio to poorer, female and other disadvantaged groups of farmers, the blueprints for measuring performance on ranches and plantations must be set aside. New tools are needed to monitor improvements on micro farms and in micro enterprises run by targeted groups. The skills of rapid rural appraisal have a larger role. Even though part of the Bank's present portfolio escapes the criticism of feeble M&E, the portfolio is shifting and pulling M&E with it.

6.18 **Relationship between KPI and M&E.** Having distinguished between sectors, one can better appreciate the problem posed by separate treatment of KPI and M&E. PMTF and Next Steps missed an opportunity to reassert the overriding importance of getting the borrowers and project managers "on board" M&E, in order to guarantee delivery of all but the easiest KPI. The linkage is mentioned in the PMTF report's annexes, but diluted in the main text. The latter says the borrowers have to be involved, but does not say how to get them there. And the Next Steps is silent on the issue: its Action Plan supports ECDP, but leaves the relationship with project M&E fuzzy. The linkage can only be secured if the KPI are set firmly in the context of the practical M&E systems that are needed to deliver them. The history of KPI in the early 1980s is a powerful reminder that KPI initiatives on their own can become impractical and collapse - because they come adrift of the institutions that must generate much of the information. The Bank should not put the cart before the horse. The primary issue for KPI is feasibility, and staff will have to respond on a country and even project specific basis regarding the capacity to collect the data. Actually what is needed is a balance. On the one hand, one should neither design nor implement an M&E system before agreeing on what information the system is supposed to collect. The KPI show where the problems are; the M&E diagnostic studies turn in that direction. On the other hand, one cannot plan for KPI without knowing how they will be collected. Where government or project management's commitment to collecting data is likely to be weak, the list of "do-able" KPI is shortened. KPI without committed capacity are non-viable.⁴² Some of the teams searching for KPI understand this and are looking for proxies for the more demanding indicators precisely in order to soften the capacity problem. But this is not true for all of the indicator papers submitted to OPR. The agriculture collection, for example, is rich with indicators grouped by subsector, but some of them have been and will continue to be almost impossible to collect.

6.19 Institution building in M&E has to respect the interests and data needs of project management, government and the donors alike, or the information generated may be sent straight to storage. The KPI most likely to receive the attention agreed at appraisal are those that are perceived by project management to be critical to successful implementation. The PMTF report's main text says "the KPI of interest to the Bank should be a subset of those indicators considered important to and sought by the project" (page 28). That is the correct posture, and KPI that do not fit that prescription and are of interest primarily to the Bank must be dealt with by special measures. They must never dominate the list of essential KPI, as the draft indicator paper for the education sector warns:

41. M&E in education has moved beyond the quality of inputs to the quality of the output: the impact of schooling on student achievement, graduation rates, preparedness for employment, etc. The problem for evaluators in this sector is to find the appropriate instruments for measuring these relationships.

42. KPI are useless if they cannot be observed and measured easily, and also if they do not respond to project stimulus.

"Mandated efforts may encourage data collection, but caution should be used in following this path which may result in a lack of "ownership", unless clear agreement on indicators is reached with Clients."⁴³

6.20 Use of Conditionality and Phasing for Investment Projects. KPI can be used to support the project by revealing problems, shortfalls and other constraints which project management, government and the Bank can then address. But KPI can also be used by the Bank aggressively to block or redirect disbursements if benchmark indicator targets have not been reached. This is normal practice with adjustment lending, covenants which are usually not applied to investment operations. Nevertheless the study found several new projects which plan to put KPI to work. One way is through making the policy reform indicators explicit in the agreed implementation schedules or even the conditionality: the best example was the set of two Brazilian education projects (Box 5). Another way, which in fact implies a major change in the project concept itself, is to tranche the investments by leading with a pilot and/or demonstration phase. In this case implementation indicators, rather than policy indicators, are targeted, and the bulk of the investment program is withheld pending evidence that the first-year targets have been reached and project design is practical.⁴⁴ There were at least two such projects in the study sample, both in Indonesia (para 3.32). Bank staff traditionally shun the aggressive use of indicators, because of the dangers of applying inappropriate KPI, the uselessness of relating disbursements to KPI that cannot be measured, or the likelihood that project managers will react by hiding the problems. This proposal worried several of the commentators on the draft of the report, who insist that the rule be limited and applied on a case by case basis. Nevertheless, there are more opportunities for aggressive use of KPI than are admitted.

B. Recommendations

6.21 Recommendations for special actions to promote M&E in the Bank are separated from those which would reinforce improvements already stimulated by the Portfolio Management Task Force and Next Steps. A short statement follows on organizational responsibilities. A few M&E design issues are then addressed, and the section ends with four suggestions for additional study.

(i) Promoting M&E in the Bank

6.22 Most Bank staff have been sensitized to the importance of M&E and need help in putting it to work. The objective is to mainstream and internalize the M&E culture that is now emerging in scattered activity by individual task managers. Most government and project staff are not yet "on board". Building an M&E culture at the country and project levels will take longer. Expectations must be tailored to the reality that M&E is a difficult, unfamiliar and sometimes unfriendly exercise, and that host-country project staff are likely to seek to minimize their inputs and hide some of the findings. KPI requirements must also be modest as well as practical. The objective is not just participation by host country project staff in a Bank-imposed M&E system. It is to build self-sustaining monitoring and evaluation habits. Better Bank M&E work, and paying attention to

43. "Indicators in Education Projects Paper." March 4, 1994, page 8. The report on poverty indicators (para 5.4) says the same in another way: "standardized indicators are likely to run counter to the emphasis on project 'ownership' and may result in simple 'box filling'."

44. A recent Bank working paper recommends displacing altogether the Bank's traditional concept of the project cycle, especially for new-style projects admitting a high degree of uncertainty about the expected course of implementation and outcome. The phasing of Bank disbursements is central to the proposal. ("Institutional Learning and Bank Operations: the New Project Cycle". Picciotto, R. Draft report dated April 19, 1994).

progress at the project level, will create their own incentives for country project staff to get involved. But the first step is for the Bank *to take M&E seriously* - to back up its exhortations with professional and institutional support.

6.23 (a) **PROVIDE INSTITUTIONAL SUPPORT IN THE CVPUS AND REGIONS.** An institutional response is warranted because of the intense interest that is certain to be maintained at senior management level in continued pursuit of the PMTF objectives. Also, as this report demonstrates, there is a discernable increase in support, activity, and quality. Both reasons provide the basis for recommending that something be done at the structural level to maintain and fortify the momentum. At present there is no system to develop, replicate and sustain the "best practice" M&E designs of the new projects, the ones put in some of the Boxes in this report but too closely associated with individual staff members. When the appraisal task manager moves on to other assignments, the successor may not have the skills to continue. Good M&E rests on a foundation of professional skills, and cannot be developed simply by recasting the Operational Directives. The weakness of the Bank's statistical support service (paras 2.15ff) is another argument for upgrading expertise through systematic training and advisory services. This should begin in the three Central Vice Presidencies, but, given the unique characteristics of M&E as well as KPI in many of the sectors, responsibility for technical support to the operational divisions should rest with individual sectoral departments. Final responsibility for promoting the appropriate use of M&E, however, rests with the Regions.

6.24 Management needs to address where and in what strength this support should materialize. Clearly, in terms of number of staff with full or part-time responsibilities for M&E, the total could be substantial if posts were to be identified in each sector department and Region. At a minimum that would mean about 15 positions, at least some of them full time. Centralization would make sense, at least up to the CVPU and Regional levels, to provide critical mass.

6.25 (b) **ESTABLISH A BANK M&E TRAINING PROGRAM.** This is part of the basic furniture for upgrading the whole of the M&E program. It allows M&E support from the CVPU's and Regions - in collaboration with the professional trainers - to be disseminated more rapidly within the operational complex. EDI's experience in training for M&E is valuable to the Bank, and its involvement in a Bank-oriented training program would be desirable. Admittedly, a revived EDI training program for M&E would serve mostly Borrower needs. But EDI involvement would help ensure that Bank and Borrower needs are both covered in a coherent way.

6.26 (c) **ENLIST SPECIAL SKILLS.** Many of the decisions on M&E are common sense, and the professionals have equipped regular staff with useful manuals. But new skills have to be brought in to any Bank-wide M&E support operation. For example, as already mentioned, statistical methods for both formal and rapid field survey and for data handling and analysis. An M&E support operation must also offer advice from persons with long experience in managing M&E, and these persons are noticeably missing from the Bank's roster.

6.27 (d) **NETWORK THE BANK.** An effective way to help develop the Bank's M&E capacity is to build upon the enthusiasm and commitment of task managers already scattered through the operational complex. The variety of M&E approaches is remarkable, and this wild flowering has occurred without prompting. The innovators can learn from each other, and the review suggests that other staff would be encouraged once exposed to progress of their colleagues. It is recommended that these energies be organized and exploited through a guided network of present volunteers and future subscribers.

6.28 (e) **FOCUS ON CERTAIN SECTORS.** Although there is a general deficiency in the use of monitoring and evaluative tools for project work, in some applications, and in some sectors, Bank practice is adequate. Focusing allows the Bank to set priorities in organizing and staffing an M&E campaign. It would give emphasis to the sectors referred to in this report as the social programs - agriculture, education, etc. These are the sectors where processes rather than technical blueprints matter, where reasonable effective MIS cannot be taken for granted.

6.29 (e) **Create a Temporary Position at the Center for Launching M&E** The Bank should consider establishing a temporary berth for a person responsible to help launch the M&E program. This person should have broad experience with indicators, surveys and project M&E systems. But the primary job would be to help ensure that appropriate technical expertise is available to the CVPUs and Regions both to build the foundations for effective M&E and to develop the training program. The assumption is that the Bank is committed to good M&E, but needs help putting it in place. The main recommendations of this section – that the CVPUs and the Regions move aggressively to establish that M&E capability, and that they be supported by a strong training program – leave room nevertheless for a catalytic agent to help set up the program rapidly and symmetrically. The person would neither clear M&E proposals nor manage M&E components. The position would wear away as the operational capability takes permanent shape. The catalytic role, and the temporary berth, belong to one of the central offices. For the longer term, after an effective M&E culture is established, OED can play a supportive role vis a vis operations. That might imply OED creating its own M&E advisory position.

6.30 This is not a call for establishing another permanent M&E support unit at the center. The fact that AGRME made no discernable difference in enhancing the sustainability of M&E in the portfolio is ample warning that unless M&E is mainstreamed it will not flourish. A special unit enhances the appearance of the separateness of M&E activity. It runs the risk of deflecting a real commitment by staff to M&E - as long as it is seen as an add-on component. And it works against the Bank's interest in internalizing the learning culture. Certainly, if support for M&E is no stronger now than it was in the 1980s, neither a temporary position nor a permanent unit will succeed. But the study finds good evidence that the commitment and momentum are genuine.

(ii) Reinforcing Next Steps

6.31 By responding forcefully to the demands of PMTF, Bank management is already moving in directions that will enhance the contributions of M&E to improving the quality of the portfolio. Among the more important of these adjustments are:

6.32 (a) **STRENGTHENING THE ROLE OF THE M&E GATEKEEPER: THE PROJECT ADVISOR.** In many departments the Project Advisor has played an important role in encouraging interested staff in developing their plans for M&E. But the contribution is unsystematic and uncertain. When the Next Step's indicative KPI lists are in place, and attention turns to M&E capacity issues, the role of the Project Advisors in reviewing both KPI and M&E design should be formalized. But they will need institutional back-up on technical matters.

6.33 (b) **DEVELOP POLICY AND OTHER RISK INDICATORS.** The area where a deficit still exists across all sectors is in tracking policy and exogenous variables influencing project performance - the risk, or enabling, indicators. Next Steps is consolidating work underway in a few sectors to define these variables. But dissemination to operational staff is still limited. Strategies for in-country capacity building to meet *these* KPI concerns differ from those aimed at project management, since the offices most likely to be involved are in the central ministries. Some of these indicators and tracking systems

can be handled directly by the Bank, for example the ones following international commodity prices. But others can only be generated with government support.

6.34 (c) **LINK THE KPI TO THE PROJECT PLANNING PROCESS.** This point repeats a PMTF recommendation that has already gained wide currency in the Bank. The logframe is one routine which forces the project planner to list the indicators which will guide subsequent supervision. The simplicity of logframe offends some practitioners, but that very quality helps sell it to less sophisticated host-country project staff who must also monitor progress. It is important that the relationship of the indicators to the objectives be transparent, and that project staff understand why the KPI help managers track progress. Logframe does both of those jobs. The fact that the logframe structure puts evaluation over monitoring carries a danger of concentrating energies on evaluation tasks with high intellectual appeal and low payoff. A proper balance must be maintained.

6.35 (d) **STRENGTHEN ICR/PAR REPORTING ON M&E, INCLUDING KPI.** The absence of information in the ex-post evaluation reports on these critical management tools was an understandable omission in an era when the Bank was not paying much attention to M&E. In the post-PMTF era, inattention is indefensible and the reporting will have to expand and be made mandatory.

6.36 (e) **IMPROVE AND EXPAND ECDP SUPPORT FOR PROJECT M&E.** The Evaluation Capacity Development Program has been mainly concerned with creating an evaluation capability at the level of the central government. The experience to date has been uneven. Many of these units have focused on monitoring the budgets of public expenditure programs, and almost no attention has been given to technical support for project M&E operations. It is proposed that ECDP expand its scope to include specialized support for the central agencies: to support, in turn, the project units. The parallel with the Bank's predicament is obvious. Project staff in both the Bank and the countries responsible for M&E are on their own, and need professional back-stopping.

6.37 (f) **CONSIDER SPECIAL SUPPORT FOR POST-COMPLETION REPORTING.** In all projects where KPI are not already incorporated in a sustained MIS process, PMTF expects that the flow of project information will nevertheless be maintained in the post-completion period. Indeed post-project reporting is central to the whole PMTF plan: in its absence the essential lessons of the operation will often be lost. ECDP can assist in encouraging governments to sponsor the work, but most countries fall outside that embrace. The Bank may have to support data collection and submission. One option is for the Bank to set up a special fund for assisting borrowers to carry out M&E subsequent to project completion. Japan's OECF has pioneered this type of arrangement with its Special Assistance for Project Sustainability program (SAPS).

6.38 (g) **INTEGRATE THE ODS.** OPR is presently recasting the OD format, and it is opportune therefore to fill the gap between the two ODs controlling Bank M&E/KPI work. The KPI sections of the Project Supervision OD 13.05 should be integrated with or at least repeated in the Project Monitoring and Evaluation OD 10.70. This will help "bind" the KPI and M&E directives. The present separation is artificial and misleading.⁴⁵

6.39 (h) **PROMOTE PARTICIPATORY EVALUATION, JUDICIOUSLY.** PMFT underlines the importance of beneficiary participation in project management, including both design and monitoring functions. Participatory evaluation is one phase of this process. It is basic to "structured learning" and an integral feature of a few of the new projects. The payoff in terms of local ownership and sustainability is expected to be high. But this is unexplored territory for the Bank. There is mixed

45. OPR is presently preparing new directives and best practice statements on these points.

evidence in the Bank on whether beneficiaries can take up the assignment without substantial guidance. The paper on *Participatory Evaluation* in water and sanitation shows that they can, and with enthusiasm. But success will be project specific, and difficult to predict. Thus the project has to be designed in a way to allow maximum room *and* support for PE.

(iii) Organizational Responsibilities

6.40 OED has not tried to assume a dominant role in promoting and supporting project M&E in the Bank and borrowing countries. It has a special contribution to make, based on its audits and studies of revealed experience in M&E and based as well on its institutional mandate to carry out the ex-post evaluative function. But for M&E to take hold in the Bank, it has to be mainstreamed - and internalized - and therefore primary responsibility for enhancing the quality and quantity of M&E rests with the Central Vice Presidencies and the Regions. This is consistent with PMFT's recommendation on ECDP.

6.41 (a) **LOCATION.** The CVPU's should be made responsible for Bank-wide, professional support to M&E, and for perfecting the KPI instruments appropriate to their sectors. However, the Regions ultimately will have primary responsibility for ensuring the M&E program functions effectively. They should consider reestablishing M&E positions in the Technical Departments, with close linkage to the new CVPU support staff. OED would also play a supportive role, both within the Bank and in marketing M&E through ECDP. The Policy Research Department has a role as well, to sponsor "evaluative research" on project impact and causality in areas beyond the reach of project M&E staff (see also para 6.45). The Training Division, supported by EDI, has perhaps the most important assignment of all.

(iv) Some Technical M&E Issues

6.42 (a) **KEEP KPI BOUND IN THE CONTEXT OF M&E.** The Bank treats KPI as a tool of good management. But KPI assessing all types of project performance variables, and KPI assessing policy and other endogenous risk factors at the sector level, depend on in-country institutional capacities to deliver them: in fact two *different* institutional systems. Also, attempts in the Bank to identify sector-specific KPI lists have a tendency to emphasize indicators of particular interest to the Bank, which weakens the resolve of ministry and project staff to deliver them. Thus, any KPI exercise should be set up within the context of the institutional capacity to deliver. Managers of the Next Steps' KPI exercise are conscious of the linkage, but instructions for the exercise have not given enough attention to feasibility and practicality. Nor should the Bank permit whatever follow-up is planned to the present indicator exercise to drift free from capacity considerations. That happened in the early 1980s during the last surge of KPI, and the movement was not sustained.

6.43 (b) **PUT KPI TO WORK, BY CONDITIONING AND PHASING DISBURSEMENTS.** In search of development impact, the Bank should make more aggressive use of the KPI, by matching disbursement schedules with the successful achievement of policy and implementation targets. The record shows that governments will accept and work to these schedules, provided the dialogue has been mutual and well handled. Phasing investments to allow for a demonstration period makes sense for many of the social programs where there is uncertainty about the feasibility or optimality of appraisal design. Conditioning is not a blanket recommendation for all or even most projects. For the majority, KPI cannot be dimensioned with unambiguous and acceptable targets, and the use of aggressive KPI conditionality can backfire.

6.44 (c) **SIMPLIFY AND SHIFT THE RESPONSIBILITIES FOR FORMAL EVALUATION.** The experience of project M&E systems with large-scale statistical surveys, and with evaluating impact on beneficiary welfare or the regional economy, has been poor. Even where expatriate consultant expertise has been recruited, the results have been unimpressive. M&E design should minimize the burden on project staff of these sophisticated studies. First, the Bank should accept that impact studies are generally better conducted outside the project and on a selective basis. Second, formal surveys of large samples and econometric methods should be used sparingly. The tools of rapid appraisal are good enough to handle most questions about outputs and impact. That does not preclude the use of formal analysis, but recognizes that the potential advantages over informal methods are often lost in practice. The Bank has not expected the infrastructure projects to measure economic impacts ex-post on a routine basis. The fact is that there has not been enough of these studies - to assess the validity of the original appraisal assumptions - even on a selective basis. But that deficit should be made up as part of the Bank's sector work or research program.

6.45 (d) **Expand the Program for Evaluative Research.** In fact, the Bank needs to do more "evaluative research" using formal experimental methods: setting up a baseline for subsequent project work and returning later to measure the attributable outputs. PRDPH's research proposal for assessing the impact of decentralization of education projects is a good example⁴⁶. That type of formal (quantitative) impact study has nothing to do with normal on-going project evaluation work and should not be imposed on the project M&E system.

(v) **Further Study**

6.46 The following are among the subjects which could not be covered during the Overview:

- (a) Close observation of the success in implementing the new, high-M&E-content projects. The poor experience of the last decade warns against taking for granted that good M&E design means effective M&E in the field.
- (b) Marshalling and assessing the evidence that better M&E has significant effects on project management and project impact, and to determine where the effects of better M&E are likely to be most evident.
- (c) The comparative advantages revealed in practice by formal and rapid field survey. There has been a trend in formal field survey in Bank programs to reduce sample size and simplify analytical technique, and this is seen as an alternative to abandoning formal, scientific method. It is important to see if the reduced format for the formal approach works, and in what ways rapid procedures are nevertheless more attractive.
- (d) Experience at the project level with participatory evaluation, to identify the circumstances where it works best.
- (e) The role of NGOs in M&E.

46. "Impact Evaluation of Decentralization and Privatization in Education Projects". Research Proposal - Draft 2. Jiménez, Emmanuel. Policy Research Department, Poverty and Human Resource Division. December 20, 1993

OED's Sample of Old and New Projects

Table 1: Distribution of Projects in the Desk Review

Sector	Africa		East Asia & Pacific		South Asia		Europe & Central Asia		Latin America		Middle East & N. Africa		Total for All Regions	
	New	Old	New	Old	New	Old	New	Old	New	Old	New	Old	New	Old
Agriculture	6	6	3	5	2	5	2	1	3	2	4	2	20	21
Education	6	6	1	2	2	0	0	0	1	2	0	1	10	11
PHN	4	6	2	3	1	0	1	0	1	1	1	1	10	11
Urban	2	3	0	2	1	2	0	0	4	3	3	0	10	10
WSS	0	0	2	1	1	0	1	0	1	3	0	2	5	6
Industry	3	1	2	2	0	0	0	1	0	0	0	1	5	5
Transport	3	4	3	1	0	0	2	1	2	3	0	1	10	10
Oil & Gas	0	0	2	0	0	2	1	0	0	0	0	1	3	3
Power	2	2	2	2	1	1	0	0	0	2	0	0	5	7
Telecom	1	2	2	0	0	1	1	0	0	1	1	1	5	5
Total	27	30	19	18	8	11	8	3	12	17	9	10	83	89

* Excluded are: adjustment projects, almost all credit projects, emergency operations and TA. Mining and environment are included in other categories.

Table 2: Completed (Old) Projects Included in the Desk Review

No.	Sector	Country	Region	Project Name	Ln./Cr. Number	Date of Approval	Amount Approved	Closing Date	Amount Disbursed	Rating	Type of Report	Report Number
1	Agriculture	Bangladesh	SA	Second Rural Development Project	C1384	14-Jun-83	100.0	30-Jun-91	82.1	S	PCR	11851
2	Agriculture	Benin	AF	Zou Province Rural Development Project	C1314	21-Dec-82	20.0	30-Sep-92	19.9	S	PCR	11887
3	Agriculture	China	EA	Forestry Development Project	C1805	11-Jun-85	47.3	31-Dec-91	42.3	S	PCR	12012
4	Agriculture	China	EA	Red Soils Area Development Project	C1733	09-Sep-86	40.0	30-Jun-92	39.9	S	PAR /2	12033
5	Agriculture	Dominican	LA	Sugar Rehabilitation Project	L1780	18-Sep-79	35.0	30-Jun-91	33.4	U	PCR	11869
6	Agriculture	Egypt	MN	Irrigation Pumping Stations Rehab.	L2270	28-Apr-83	41.5	30-Jun-92	40.5	S	PCR	11870
7	Agriculture	Hungary	EC	Crop Production Improvement Project	L2738	08-Jul-86	100.0	31-Dec-91	100.0	S	PCR	11888
8	Agriculture	India	SA	Karnataka Social Forestry Project	C1432	20-Dec-83	27.0	31-Mar-92	24.7	S	PCR	11929
9	Agriculture	India	SA	Nabard Credit Project	L2853	25-Feb-88	750.0	30-Jun-91	750.0	U	PAR	12109
10	Agriculture	Indonesia	EA	Yogyakarta Rural Development Project	C0848	12-Jul-79	24.0	31-Mar-89	22.3	U	PAR	11981
11	Agriculture	Pakistan	SA	Integrated Hill Farming Development	C1481	17-Apr-84	21.0	31-Mar-92	14.3	S	PCR	12032
12	Agriculture	Paraguay	LA	Seventh Livestock Development Project	L2372 /1	03-Jan-84	32.1	31-Dec-90	23.7	S	PCR	11838
13	Agriculture	Senegal	AF	Agricultural Research Project	C1178	08-Sep-81	18.5	31-Dec-89	18.5	S	PCR	11836
14	Agriculture	Solomon Isl.	EA	Rural Services Project	C1430	20-Dec-83	3.5	31-Dec-90	2.2	U	PCR	11930
15	Agriculture	Somalia	AF	Northwest Region Agricultural Devt.	C1538	08-Jan-85	10.8	30-Jun-91	10.8	U	PCR	11919
16	Agriculture	Sri Lanka	SA	Fourth Tree Crops Project	C1562	21-Mar-85	56.0	31-Dec-91	55.0	S	PCR	12045
17	Agriculture	Thailand	EA	Land Tilling Project	L2440	12-Jun-84	35.0	31-Mar-92	33.8	S	PCR	11898
18	Agriculture	Togo	AF	2nd Rural Devt. Proj. In Cotton Areas	C1302	23-Nov-82	60.8	31-Mar-91	47.0	S	PCR	11839
19	Agriculture	Tunisia	MN	Northwest Rural Development Project	L1997	19-May-81	48.0	02-Jan-90	32.9	U	PAR	12031
20	Agriculture	Zaire	AF	Seeds Project	C1809	13-Jun-85	14.9	30-Jun-92	9.2	U	PCR	11897
21	Agriculture	Zimbabwe	AF	National Agric. Extension And Research	L2335	07-Jul-83	13.1	30-Sep-91	10.8	S	PCR	11835
22	Education	Botswana	AF	Fourth Education Project	L2844	17-Dec-85	28.0	30-Jun-91	23.7	S	PCR	11184
23	Education	Brazil	LA	Urban Basic Ed. For North & Center-West	L2412	17-May-84	40.0	28-Feb-91	30.8	S	PCR	11889
24	Education	Burundi	AF	Third Education Project	C1368	10-May-83	15.8	30-Jun-88	15.8	S	PAR	10982
25	Education	Djibouti	AF	First Education Project	C1543	22-Jan-85	5.0	30-Jun-91	4.9	S	PCR	11284
26	Education	Ecuador	LA	Second Vocational Training Project	L2171	03-Jun-82	18.0	30-Jun-90	15.8	S	PCR	11389
27	Education	Jordan	MN	Fifth Education Project	L2248	15-Mar-83	37.8	30-Jun-88	28.5	U	PAR	11513
28	Education	Lesotho	AF	Fourth Education Project	C1512	31-Jul-84	10.0	30-Sep-91	10.0	S	PCR	11182
29	Education	Mali	AF	Third Education Project	C1442 /1	01-Mar-84	9.5	30-Sep-90	8.5	S	PCR	10985
30	Education	PNG	EA	Secondary Education Project	L2395	27-Mar-84	48.3	31-Dec-90	48.3	S	PCR	11472
31	Education	Philippines	EA	Vocational Training Project	L2200	21-Sep-82	24.4	30-Apr-91	14.9	S	PCR	11163
32	Education	Zaire	AF	Education Technical Asstce & Training	C1519	11-Sep-84	9.0	31-Mar-91	8.1	U	PCR	11283
33	PHN	Botswana	AF	Family Health Project	L2413	17-May-84	11.0	31-Jan-92	11.0	S	PCR	12014
34	PHN	Brazil	LA	National Health Policy Studies Project	L2448	21-Jun-84	2.0	31-Dec-89	1.7	U	PCR	11503
35	PHN	China	EA	Rural Health And Medical Education	C1472	08-May-84	85.0	31-Dec-91	85.0	S	PCR	12048
36	PHN	Comoros	AF	Health And Population Project	C1408	23-Aug-83	2.9	30-Jun-91	2.8	U	PCR	11257
37	PHN	Ghana	AF	Health And Education Rehabilitation	C1653	23-Jan-86	15.0	31-Dec-91	15.0	S	PCR	11080
38	PHN	Guinea-Bissau	AF	Population, Health And Nutrition	C1800	21-May-87	4.2	31-Dec-91	4.2	U	PCR	11759
39	PHN	Indonesia	EA	Second Nutrition And Community Health	L2636	28-Nov-85	33.4	31-Mar-92	31.8	S	PCR	11997
40	PHN	Kenya	AF	Integ. Rural Health And Family Planning	C1238	04-May-82	23.0	31-Dec-90	22.2	S	PCR	11079
41	PHN	Mali	AF	Health Development Project	C1422	08-Dec-83	18.7	30-Sep-91	18.4	U	PCR	11502
42	PHN	Philippines	EA	Second Population Project	C0923	18-May-79	40.0	30-Jun-88	32.3	U	PAR	09380
43	PHN	Yemen	MN	Health Development Project	C1377	31-May-83	7.8	31-Dec-88	7.6	S	PCR	09928
44	Urban	Brazil	LA	Urban Transport 3	L1985	31-Mar-81	80.0	31-Dec-86	80.0	S	PAR	10832
45	Urban	Colombia	LA	Earthquake Reconstruction	L2379	02-Feb-84	40.0	30-Jun-88	40.0	S	PCR	10704
46	Urban	Ethiopia	AF	Urban Development	C1399	17-May-83	20.0	30-Jun-91	20.0	S	PCR	10721
47	Urban	Guinea	AF	Conakry Urban	C1489	01-May-84	15.2	30-Jun-91	15.2	S	PCR	10722
48	Urban	Haiti	LA	Urban Development Project	C1334 /1	22-Mar-83	21.0	31-Dec-91	17.8	U	PCR	11481
49	Urban	India	SA	Madhya Pradesh Urban Development	L2329	28-Jun-83	24.1	30-Jun-91	12.5	S	PCR	11447
50	Urban	India	SA	Housing Development Finance Corp. Project	L2929	31-Mar-88	500.0	30-Sep-91	500.0	S	PCR	11483

No. Sector	Country	Region	Project Name	Ln/Cr. Number	Date of Approval	Amount Approved	Closing Date	Amount Disbursed	Rating	Type of Report	Report Number
51 Urban	Indonesia	EA	Fifth Urban Development Project	L2406	08-May-84	38.3	30-Jun-91	38.9	S	PCR	11501
52 Urban	Kenya	AF	Secondary Towns	C1360 /1	18-Jun-83	29.0	31-Dec-90	22.0	U	PCR	10748
53 Urban	Korea, Rep.	EA	Urban Land Development Project	L2704	29-May-86	150.0	31-Dec-90	150.0	S	PCR	11519
54 Water Supply	Brazil	LA	Water Supply And Sewerage Project	L2249	22-Mar-83	302.3	30-Sep-88	302.3	U	PCR	11531
55 Water Supply	Brazil	LA	Tech. Asstcs & Pilot Rural Water Supply	L2532	07-May-85	16.3	30-Sep-90	11.1	U	PCR	11549
56 Water Supply	Indonesia	EA	Jakarta Sewerage And Sanitation Project	L2236	08-Feb-83	22.4	31-Mar-91	21.8	S	PCR	12150
57 Water Supply	Jordan	MN	Eight Cities Water Supply And Sewerage	L2425	24-May-84	60.0	30-Jun-90	60.0	S	PAR	11608
58 Water Supply	Mexico	LA	Third Medium-Size Cities Water/Sewerage	L2281	17-May-83	200.6	31-Dec-90	180.6	S	PAR /2	12145
59 Water Supply	Monroco	MN	Third Water Supply Project	L2008	28-May-81	87.0	31-Dec-88	78.0	S	PCR	11861
60 Industry	China	EA	Fertilizer Rehabilitation	L2541	14-May-85	97.0	31-Dec-90	98.3	S	PCR	10823
61 Industry	Jordan	MN	Second Arab Potash Project	L2786	10-Mar-87	24.0	30-Jun-92	24.0	S	PCR	11290
62 Industry	Myanmar	EA	Textile Finishing Plant Project	C1425	13-Dec-83	29.7	31-Dec-90	28.5	U	PCR	11580
63 Industry	Portugal	EC	Mechanical Industries Project	L1875 /1	19-Jun-80	88.0	31-Dec-83	84.0	U	PAR	12071
64 Industry	Tanzania	AF	Mufindi Pulp And Paper Project	C0875 /1	04-Jan-79	60.0	31-Dec-82	60.0	U	PAR	11481
65 Transport	Brazil	LA	Federal Highways Sector Project	L2446	21-Jun-84	210.0	31-Dec-88	204.8	S	PCR	11082
66 Transport	Egypt	MN	Port Sald Expansion And Rehabilitation	L2569	08-Jun-85	37.0	30-Jun-91	17.3	S	PCR	11033
67 Transport	Gambia	AF	Banjul Port 2	C1266	17-Jun-82	6.5	31-Mar-91	8.5	S	PCR	10778
68 Transport	Hungary	EC	Transport (Rail/Road)	L2577	28-May-85	75.0	30-Jun-91	75.0	S	PCR	10778
69 Transport	Jamaica	LA	Highway Maintenance Project	L2263	26-May-83	15.0	30-Jun-89	15.0	U	PCR	11128
70 Transport	Korea, Rep.	EA	Railway 7	L1836	29-Apr-80	94.0	31-Dec-84	93.0	S	PAR	10777
71 Transport	Lesotho	AF	Fourth Highway Project	C1481	24-May-84	15.2	31-Aug-91	15.1	S	PCR	10691
72 Transport	Mexico	LA	Railway 4	L1829	16-Dec-80	150.0	30-Jun-86	149.9	S	PAR	10817
73 Transport	Seychelles	AF	Transport	L2462	12-Feb-85	6.2	31-Dec-90	6.2	S	PCR	10786
74 Transport	Sudan	AF	Second Port Project	C1233	20-Apr-82	25.0	30-Jun-88	23.8	U	PCR	11004
75 Energy	Bangladesh	SA	Energy Efficiency And Refinery Rehab.	C1357	10-May-83	57.0	31-Dec-88	28.1	S	PAR	12120
76 Energy	India	SA	Offshore Gas Development	L2241	24-Feb-83	222.3	31-Dec-88	135.7	S	PCR	09092
77 Energy	Jordan	MN	Energy Development Project	L2371	20-Dec-83	90.0	31-Dec-90	58.8	S	PAR	11785
78 Power	Bangladesh	SA	Rural Electrification Project	C1262 /1	15-Jun-82	46.0	30-Jun-90	47.9	S	PAR /2	12111
79 Power	Brazil	LA	Rural Electrification Project	L2365	08-Dec-83	222.8	30-Jun-90	222.0	U	PCR	11529
80 Power	Burundi	AF	Transmission And Distribution Project	C1563	09-May-85	24.6	31-Dec-92	18.9	S	PAR /2	12157
81 Power	Colombia	LA	Village Electrification Project	L1998	21-May-81	36.0	30-Jun-88	30.1	S	PCR	12039
82 Power	Indonesia	EA	Fourteenth Power Project	L2443	14-Jun-84	210.0	30-Sep-91	205.1	S	PCR	11014
83 Power	Thailand	EA	Accelerated Rural Elec. 2	L1871	12-Jun-80	75.0	30-Jun-87	75.0	S	PAR	10856
84 Power	Zimbabwe	AF	First Power Project	L2212	07-Dec-82	105.0	30-Jun-90	105.0	S	PAR /2	12156
85 Telecom.	Bangladesh	SA	Telecommunications 3	C1321	25-Jan-83	36.0	30-Jun-91	29.4	S	PCR	10644
86 Telecom.	Burkina Faso	AF	Third Telecommunications Project	C1235	27-Apr-82	17.0	30-Jun-86	17.0	S	PCR	09167
87 Telecom.	Oman	MN	Telecommunications 2	L2513	11-Apr-85	23.0	30-Jun-91	23.0	S	PCR	10743
88 Telecom.	Uganda	AF	Posts And Telecommunications Rehab.	C1367	17-May-83	22.0	30-Jun-87	22.0	U	PCR	10148
89 Telecom.	Uruguay	LA	Telecommunications 1	L2033	07-Jul-81	40.0	30-Jun-88	21.7	S	PCR	10678

/1 This project was funded with more than one loan or credit. The Amount Approved includes the total for all loans/credits for this project.
/2 In addition to the PAR listed here, there is a separate PCR for the project.

Table 3: Recently Approved (New) Projects in the Desk Review

No.	Sector	Country	Region	Project Name	L/C#	Number	Date of Approval	Amount Approved
1	Agriculture	Albania	EC	Rural Poverty Alleviation	C24810		23-Feb-03	2.4
2	Agriculture	Burundi	AF	Agri-Business Promotion	C24180		03-Sep-02	3.1
3	Agriculture	Chile	LA	Irrigation Development	L26280		17-Nov-02	46.0
4	Agriculture	China	EA	Grain Distribution Project	C28180		17-Jun-03	168.0
5	Agriculture	Colombia	LA	Agricultural Credit	L26750		18-Mar-03	250.0
6	Agriculture	Egypt	MN	Marsh Resources Management	C25040		27-May-03	22.0
7	Agriculture	Ghana	AF	Environment	C24280		15-Oct-02	18.1
8	Agriculture	Ghana	AF	Livestock	C24410		08-Dec-02	22.5
9	Agriculture	Hungary	EC	Product Market Development	L26090		13-Jul-02	100.0
10	Agriculture	India	SA	UP Soild Lands Reclamation	C28100		10-Jun-03	54.7
11	Agriculture	Indonesia	EA	Groundwater Development	L26880		13-Apr-03	84.0
12	Agriculture	Iran	MN	Irrigation Improvement	L26700		16-Mar-03	157.0
13	Agriculture	Kenya	AF	Drought Recovery	C24800		23-Feb-03	20.0
14	Agriculture	Nepal	SA	Sunari Morang Headquarters	C24300		05-Nov-02	28.0
15	Agriculture	Philippines	EA	Irrigation Operation Supp II	L26070		20-May-03	51.3
16	Agriculture	Tunisia	MN	Second Forestry Development	L26010		11-May-03	69.0
17	Agriculture	Turkey	MN	E. Anatolia Watershed Rehab.	L26870		11-Mar-03	77.0
18	Agriculture	Uganda	AF	Agri. Research & Training	C24480		15-Dec-02	26.0
19	Agriculture	Venezuela	LA	National Parks Management	L26290		06-Dec-02	56.0
20	Agriculture	Zambia	AF	Marketing & Processing	C24220		06-Sep-02	33.0
21	Education	Brazil	LA	North East Education II	L26040		13-May-03	212.0
22	Education	Chad	AF	Basic Education	C25010		25-May-03	19.3
23	Education	Cote D'Ivoire	AF	HRD Management Support	C25050		01-Jun-03	6.7
24	Education	Ghana	AF	Primary School Development	C25080		10-Jun-03	66.1
25	Education	India	SA	Urban Pradhan Basic Education	C25090		10-Jun-03	166.0
26	Education	Mauritania	AF	Technical/Vocational Training	C25210		22-Jun-03	12.5
27	Education	Nigeria	AF	Development Communication	C24810		13-Apr-03	8.0
28	Education	Pakistan	SA	Bauzhitkan Primary Education	C24820		13-Apr-03	108.0
29	Education	Solomon Isl.	EA	Education III	C25000		26-May-03	16.9
30	Education	Uganda	AF	Primary Educ. & Teacher Dev.	C24830		13-May-03	52.6
31	PHN	Angola	AF	Health	C24900		11-May-03	18.9
32	PHN	Burundi	AF	Social Action	C24940		13-May-03	10.4
33	PHN	Colombia	LA	Municipal Health Services	L26150		06-Jun-03	50.0
34	PHN	Hungary	EC	Health Services And Management	L26870		20-Apr-03	91.0
35	PHN	India	SA	National Leprosy Elimination	C25290		26-Jun-03	66.0
36	PHN	Mozambique	AF	Food Security	C24870		27-Apr-03	6.3
37	PHN	PNG	EA	Population Project	L26810		13-Apr-03	6.9
38	PHN	Philippines	EA	Urban Health & Nutrition	C25090		06-Jun-03	70.0
39	PHN	Yemen	MN	Family Health	C25250		24-Jun-03	28.6
40	PHN	Zimbabwe	AF	STI	C25160		17-Jun-03	64.5
41	Urban	Algeria	MN	Housing Completion & Sect. Dev.	L26810		04-Mar-03	200.0
42	Urban	Brazil	LA	Metropolitan Transport - Rio	L26330		28-Jun-03	128.5
43	Urban	Mauritania	AF	Construction Capacity & Empl.	C24690		02-Feb-03	12.0
44	Urban	Mexico	LA	Transport Air Pollution Control	L26430		15-Dec-02	220.0
45	Urban	Mexico	LA	Medium Speed Cities Transport	L26890		16-Feb-03	200.0
46	Urban	Morocco	MN	Land Development	L26180 /1		10-Jun-03	130.0
47	Urban	Sierra Leone	AF	Freeway Infrastructure	C25110		10-Jun-03	26.0
48	Urban	Sri Lanka	SA	Colombo Urban Transport	C24690		18-May-03	20.0
49	Urban	Tunisia	MN	Municipal Sector Investment	L26070		02-Jul-02	73.0
50	Urban	Venezuela	LA	Low-Income Barrios Improvement	L24690		23-Jun-02	40.0
51	Water Supply	Brazil	LA	Water Quality/Plan (SP/Parana)	L26030 /1		02-Jul-02	246.0

No.	Sector	Country	Region	Project Name	Lt./Cr. Number	Date of Approval	Amount Approved
82	Water Supply	India	SA	Karnataka WS & Env/Sanitation	C24830	20-Apr-93	92.0
83	Water Supply	Indonesia	EA	Water & San. For Low Inc. Comm.	L36260	24-Jun-93	80.0
84	Water Supply	Korea, Rep.	EA	Seoul & Kwangju Sewerage	L36900	19-Apr-93	110.0
85	Water Supply	Turkey	EC	Bursa Water & Sanitation Proj.	L36660 /1	11-Mar-93	129.5
86	Industry	Angola	AF	Financial Institutions Devt.	C24210	03-Sep-92	21.0
87	Industry	China	EA	Cement	L34430	17-Mar-92	82.7
88	Industry	China	EA	So. Jiangsu Environ. Protection	L35820	25-Mar-93	250.0
89	Industry	Madagascar	AF	Financial Institutions Devt.	C24870	20-May-93	6.3
90	Industry	Mali	AF	Mining Capacity	C23900	17-Jun-92	6.0
91	Transport	Albania	EC	Transport/Infrastructure	C24900	20-May-93	18.0
92	Transport	Angola	AF	Transport Recovery	C24500	03-Sep-92	41.0
93	Transport	Argentina	LA	Road Maint. & Rehab. Sector	L38110	03-Jun-93	340.0
94	Transport	Chad	AF	Transport Sect II	C25200	22-Jun-93	37.0
95	Transport	China	EA	Shanghai Port Rest. & Devt.	L36520	22-Dec-92	150.0
96	Transport	China	EA	Railway VI	L36810	26-Mar-93	420.0
97	Transport	Ghana	AF	Urban Transport	C24890	25-May-93	76.2
98	Transport	Indonesia	EA	E. Indonesia Kabupaten Roads	L35790	23-Mar-93	155.0
99	Transport	Mexico	LA	Highway Rehab. & Safety	L36280	24-Jun-93	490.0
70	Transport	Romania	EC	Transport	L36630	13-Apr-93	120.0
71	Energy	Korea, Rep.	EA	Petroleum Dist. & Sec. Mgmt. Impr.	L38130	03-Jun-93	120.0
72	Energy	Russia	EC	Oil Rehabilitation	L38220 /1	17-Jun-93	810.0
73	Energy	Thailand	EA	Gas Transmission	L35090	07-Jul-92	105.0
74	Power	China	EA	Tianhuangping Hydro	L36090	18-May-93	300.0
75	Power	Ghana	AF	National Electrification	C24870	04-Mar-93	90.0
76	Power	India	SA	NTPC Power Generation I	L36320	29-Jun-93	400.0
77	Power	Indonesia	EA	Cirata Hydro Phase II	L36020	11-May-93	104.0
78	Power	Tanzania	AF	Power VI	C24890	06-May-93	200.0
79	Telecom.	Bulgaria	EC	Telecommunications	L36920	13-Apr-93	30.0
80	Telecom.	Indonesia	EA	Telecommunications IV	L34820	02-Jun-92	375.0
81	Telecom.	Morocco	MN	Telecom. Restructuring	L35770	02-Feb-93	100.0
82	Telecom.	Philippines	EA	Telephone System Expansion	L36220	06-Oct-92	134.0
83	Telecom.	Tanzania	AF	Telecom III	C24890	27-Apr-93	74.5

/1 This project was funded with more than one or credit. The Amount Approved includes the total for all loans/credits for this project.

Journal of International Development: Vol. 4, No. 5, 497-510 (1992)
Reproduced by permission of John Wiley and Sons Limited.

MONITORING AND EVALUATION IN AGRICULTURAL AND RURAL DEVELOPMENT PROJECTS: LESSONS AND LEARNING*

GILROY COLEMAN

School of Development Studies, University of East Anglia

Abstract: Project M&E systems are a requirement of most major funding agencies and result from a concern that many projects fail because they are badly managed. However, M&E systems themselves have a poor record. This paper argues (a) that project design is a significant factor in project failure and (b) that the over-ambitious terms of reference for M&E systems contribute to their failure. Technical issues may permit managers to ignore M&E information, but these issues may not be as important as the potential threat which information may represent. Progress is being made in addressing these issues but M&E has been slow to learn from its own experience.

1 INTRODUCTION

In recent years there has been a tremendous increase in the extent to which externally financed rural and agricultural development projects in less developed countries have been subject to formal monitoring and evaluation efforts. Methodological guidelines for project monitoring and evaluation, and project reporting procedures, exist for almost all major development agencies.

The emergence of large-scale M&E efforts has arisen largely from a concern that many early agricultural and rural development projects were failing to achieve their planned objectives. For the period up to 1968 it was noted that 'there were continuing, widespread, and often near-fatal weaknesses in the implementation of these [agricultural development] policies in the field' (Hunter *et al.*, 1976: 9). Amongst World Bank-funded projects the Director of the World Bank's Agricultural and Rural Development Department noted that 'the most widespread problem is management' (Yudeiman, 1976: 26).

* This paper was presented at the Development Studies Association Conference in Swansea, September 1991.

498 *G. Coleman*

It was as a response to these concerns that the major agencies began to introduce formal monitoring and evaluation components into agricultural and rural development projects. It was envisaged that M&E would provide frequent and quantitative checks on implementation performance (budget disbursement, achievement of physical targets, adherence to planned schedules) and that project management would be improved by the provision of information which would rapidly pinpoint bottle-necks, gaps in performance, underachievement of targets and unexpected consequences of project activities. With feedback of this sort it was hoped that management would be able to react in a positive way to eliminate project shortcomings during implementation.

During the 1970s and 1980s agricultural and rural development projects have continued to fail. Based on 192 audited rural development (RD) projects, and taking a 10 per cent economic rate of return (ERR) at completion as the cut-off point, the World Bank's Operations Evaluation Department reported (1988: table 3, p. 21) that 37 per cent of poverty-focused RD projects, and 21 per cent of non-poverty projects, failed. For RD area projects the failure rate on this criterion was 51 per cent (p. 25), including an 80 per cent failure rate for area development projects in eastern and southern Africa. There are clearly doubts about the appropriateness of equating the ERR with project success in the wider sense. However, even by other criteria many projects have failed: for 112 audited projects with food production goals about 48 per cent failed to achieve these targets at completion.

There are a number of apparent causes for this perceived failure by monitoring and evaluation to 'cure' the problems of project implementation. These are considered in two main sections below. The first main section involves an examination of the 'management failure' view.

2 PROJECT FAILURE: A MANAGEMENT PROBLEM?

One view is that the original diagnosis of the problem was wrong, i.e. that it was not entirely a problem of 'management' which could be remedied by better information provision.

One alternative diagnosis is that poor project design made many projects unimplementable, i.e. doomed to failure whatever the quality of the management. Many of the very early agricultural projects were clearly badly designed. The Groundnut Scheme in colonial Tanganyika in the late 1940s was implemented without even the most basic examination of soil and weather conditions. The Niger Agricultural Project in Nigeria in the mid-1950s provided settlers with large farms without reference to the labour requirements of local agriculture. One result was that farm families — typically consisting of a man and his wife — were required to undertake more than 400 days of work in less than 3 months without hiring labour (Baldwin, 1957: chap. 9).

Project design has certainly improved since these early days, but it remains the case that serious design faults are still built into many projects. A recent World Bank livestock project in Somalia was designed to increase offtake from herds operated by nomadic pastoralists by increasing the prices paid for animals. The pastoralists reacted rationally to this policy by selling *less* animals: their cash needs were limited by their nomadic lifestyle and their wealth was measured by the number of animals which they kept. The project allowed them to increase their wealth, by increasing their herd sizes, by selling fewer animals.

Monitoring and Evaluation in Agricultural Projects 499

More commonly, design problems can be seen in terms of unrealistic targets for activities which depend on long and cumbersome procurement procedures. These translate rapidly into substantial performance shortfalls, especially in activities such as road, bridge and dam building which depend on imported equipment. Despite the reputation of these sorts of delays in project after project the designers often seem slow in drawing the appropriate lessons and providing realistic targets based on likely procurement performance.

The problem which these examples present for project management is that it is they — management who will be blamed for poor performance when the project falls behind schedule, and attention will be focused on remedies to improve this performance rather than on remedies to improve project design.

A second area of potential mis-diagnosis of the problem is where the project appears to be well designed and well implemented but fails as a result of an unsupportive policy environment. This problem has been apparent for some time: Lele (1975: 176) noted in the mid-1970s that 'Frequently, despite the fact that the likely impact of domestic policies and institutions was anticipated. . . national policies could not be changed to improve project performance'. It is only in more recent years that more attention has been given to the policy environment in which projects are located, in order to ensure that this environment supports rather than undermines project activities.

Examples of the latter are widespread. Many relate to pricing and fiscal policies at the national level. In Nigeria:

The production history of cotton under the Funtua ADP illustrates well the inutility of extension efforts aimed at a crop for which returns are inadequate. Cotton is unique among major crops in Funtua in that almost all the production has to be sold through formal marketing channels, at a price which is (nominally at least) subject to government regulation. The controlled price remained constant throughout the Project life, at a time when prices of all other crops were moving rapidly upwards. . . Consequently there was no increase in adoption of cash inputs such as fertilizer and insecticide sprays, and the number of growers fell sharply, from 70 per cent of households in 1976/77 to only 30 per cent in 1979/80 (APMEPU, 1982: 43).

In Funtua the farmers simply switched out of cotton and into these competing crops. At the time the Cotton Marketing Board made strenuous attacks on both the design and implementation of the project.

Another example involved a project designed to increase the production and export of palm oil. The over-valuation of the local currency meant that the project could not find an export market for the oil it procured and could only find a domestic market if it sold the oil at less than the producer price. Blame for this failure was attached to what was said to be an over-expensive oil collection and processing system, but it was clear — with hindsight — that the project could not possibly succeed with the then current exchange rate policy.

In both of these projects attention was focused on implementation, and — to a lesser extent — design, and their subsequent failure was seen to be at least partly attributable to poor M&E.

A third 'cause' of M&E failure relates to the planning system itself and the extent to which it is possible for project managers to make changes in the light of information

500 *G. Coleman*

made available to them. Some planning systems remain geared to the production of rigid 'blueprint' plans in which the potential for subsequent adjustment during implementation is extremely limited.

Even if the plan itself is not rigid many bureaucratic structures are highly centralized, with decision-making ability concentrated at the top of the hierarchy. For example,

All of the South Asian countries have a considerable degree of central economic planning. Consequently, the central planning or finance agencies have often sought to use monitoring as an instrument of financial control and budget allocation. This has frequently led to protest by the line ministries that their autonomy and financial independence is being eroded by a super ministry (Ahmed and Bamberger, 1989: 4).

It is often the case that even where limited decision-making authority exists at lower levels the junior officers are reluctant to make decisions without the authority of their immediate senior. At the extreme this may mean that even relatively minor decisions end up — months later — on the Minister's desk.

The causes of project failure which we have examined so far look rather like a list of excuses for M&E: it is not really the fault of M&E systems but rather the fault of the project design, policy environment or planning/decision-making system. While this may be true up to a point we also need to look at performance of M&E systems and the reasons why their activities have apparently failed to improve project implementation. This is examined in the section below.

3 THE EXPERIENCE OF M&E

Despite more than two decades of practice the results of monitoring and evaluation of agricultural and rural development projects have not been encouraging. In the OED review (World Bank, 1988: 101) it was noted that of 104 World Bank projects with built-in M&E components 'only 15% showed good M&E results, 39% had seriously deficient M&E systems and in 46% the M&E system was either not implemented or performance was unsatisfactory'. In a similar review of IFAD projects it was noted that of 127 project M&E systems '18 per cent were rated highly effective, 27 per cent moderate, 44 per cent low and 11 per cent nil' (Coleman, 1989e: 27). This low quality of M&E performance is particularly notable since both the World Bank and IFAD have emphasized their commitment to project M&E and both organizations have committed significant funds to project M&E: to the end of 1987 IFAD financed 54 per cent of the M&E costs of its projects, amounting to more than \$25 million, and the World Bank estimates that more than \$40 million was spent on project M&E just in Nigeria between 1975 and 1983. Given this level of expenditure, and the very limited overall success of the systems, it is perhaps not surprising that van de Laar (1980: 205) noted that a group of participants in a World Bank-sponsored seminar on M&E felt that 'the Bank does not get its money's worth from [M&E] units'.

The volume of case studies edited by Clayton and Pétry (1983) provides an excellent guide to the variety of conceptual, organizational, operational and statistical problems which have hindered both the collection of appropriate information and the use of that information for management purposes.

Monitoring and Evaluation in Agricultural Projects 501

In relation to the centralized evaluation unit serving the World Bank agricultural projects in northern Nigeria, Forrest (1981: 245) noted that 'contrary to some expectations, information provided by the unit has had little or no effect on the policies of project managers'. The OED review (World Bank, 1988: 79-80) goes further: 'Because of the magnitudes involved, the Nigerian experience may well constitute the Bank's biggest disappointment with M&E' and 'The central M&E system has had little impact on ADP [Agricultural Development Project] management decisions, and the accumulated data base is not only weak but in some cases useless'.

4 M&E: SYSTEMS FAILURES

In a recent review of M&E units in IFAD projects Coleman (1990: 149) identified a series of problem areas as follows:

These problems cover the establishment of M&E units, various aspects of M&E design (specification of resources, the adequacy of financial allocations, the specifications of design in appraisal/preparation reports, over-ambitious designs, baseline surveys), staffing, and a series of issues related to M&E performance (organisational location, M&E as separate activities, beneficiary contact monitoring, project follow-up/supervision, use of M&E Information).

In a similar review of World Bank experience it was noted that:

Shortcomings are numerous and include...over-ambitious objectives and targets; inadequate provision made for M&E at project appraisal; design problems in general; organisational problems; inadequate Bank supervision, late start-up and failure to provide timely feedback information; staff shortages and failure to provide training; underfunding; management deficiencies; diversion of staff time to other duties; better coverage of engineering/implementation aspects compared with agricultural/operational matters; conflicts between project authorities and a central M&E unit; insufficient borrower interest or loss of borrower support during implementation; too much data generated by complicated surveys combined with inadequate data processing capacity; poor cooperation of farmers or participating agencies due to lack of understanding of the purpose of M&E; failure by project management to use monitoring information; and a tendency to respond inadequately after weaknesses in M&E systems had been identified (World Bank, 1988: 102).

The similarity in these listings is striking, and suggests a common set of problems affecting a large proportion of agricultural/rural development projects. There is not the space to look at all of these issues in detail (and many are covered in the above papers, plus Coleman's 1989e report) but it is worth examining several which remain both current and significant.

Over-ambitious M&E Designs

For one of the earlier World Bank projects in Nigeria part of the terms of reference were specified as follows:

Evaluation must include such standard economic criteria as changes in gross and

502 *G. Coleman*

net incomes of participants, changes in volume and value of production, and measures of efficiency in use of project resources. It must include such direct and indirect benefits as changes in the efficiency and effectiveness of the marketing system, changes in employment, changes in the volume and type of business in the region, changes in tax revenues and expenditures, and change in other economic and social indicators that would show direct and indirect results of project activities. Equally important may be changes in patterns of consumption, in participation in education, in health, and in other measures of well-being of people in the project area (World Bank, 1974: Annex 9, pp. 1-2).

These are staggering objectives and, as Slade (1983: 186) has noted, they 'reflect the temptation to include everything, instead of attempting the more difficult task of creating specific objectives, defining an analytical framework, specifying data priorities and deciding what, on balance, can be left out'. Hesling (1984) has noted that by 1981 there were seven ADPs—with similar M&E terms of reference to those above—and by 1986 this had increased to 15, involving virtually every state in Nigeria.

Many implications followed from these TOR, but before examining these it is worthwhile emphasizing that such excessive M&E requirements were by no means unique. For an Ethiopian project Coleman (1989: 15) has noted that

a relatively small M&E unit was required to collect, process and analyse data for a series of loan disbursement/repayment formats, office operational performance indicators and ongoing evaluation formats, as well as organising a set of impact evaluation studies. In practice, less than half of this workload was achieved.

Similarly, in Sierra Leone (Coleman, 1989e: 15) an M&E design mission

drew up a five-year work programme which consisted of (a) a baseline survey to establish cropping patterns, farmer's income, consumption habits, farm technology, farm inputs usage, agricultural institutions, farmer size and farm labour availability etc., (b) a series of follow-up surveys involving 100 farmers in year 1 rising to 600 in year 5, (c) a series of special studies on credit, marketing, swamp development etc. and (d) monitoring reports, price and output data reports and annual reports. A six page list of the variables and indicators to be measured and a schedule of all the studies to be done over the five years was prepared. This enormous workload was not achieved.

Coleman (1989e) quotes further evidence of similar problems in a variety of countries.

These TOR are project design faults, similar to those noted earlier, and they have the same result: when the targets are not achieved then project management (or at least the M&E unit within project management) is blamed. Thus in relation to the M&E of the Nigerian projects the OED review (World Bank, 1988, pp. 80) noted that 'management problems have been identified as the main factor explaining poor performance' and in relation to the centralized M&E unit 'APMEPU's performance has been weak'. Perhaps so, but the task was impossible anyway.

A variety of implications follow from the type of overblown M&E TOR noted above. These are examined separately below.

Monitoring and Evaluation in Agricultural Projects 503**The Focus on Evaluation**

In recent years there has been growing criticism that M&E systems, and the units which implement them, have concentrated on evaluation activities at the expense of monitoring activities. To the extent that it is *monitoring* activities which serve immediate project management needs then it appears that the 'solution' to the management problem which was supposed to be represented by M&E systems has been off-target. Indeed, it could be said that evaluation-focused M&E systems represented a cure for which there was no disease.

The need for M&E systems which served managers was articulated at a World Bank workshop in 1979: 'There was a strong consensus among the participants [of the workshop] in identifying a hierarchy of users, ranking project-level needs highest in priority, national- (or regional-) level users intermediately, and donor agencies lowest' (Deboeck and Kinsey, 1980: 7) and more recently emphasized by Casley and Kumar (1987: 21):

first priority [should] be given to the monitoring needs of project staff working at various levels of responsibility in project implementation. They should be treated as the primary information consumers. Next come provincial or national government agencies and departments that are responsible for supervising and coordinating projects and meeting the requirements of intended beneficiaries. Then come the demands of funding agencies.

In practice, however, the ambitious TOR provide an almost exclusive focus on evaluation. Casley and Kumar (1987: 10) quote a case study (clearly identifiable to insiders as APMEPU in Nigeria) which 'tended to focus on the data needs for evaluating the national food plan ... at the expense of [project] internal monitoring requirements'. This 'greater emphasis on evaluation' (Slade, 1983: 187) flowed directly from the job specification which APMEPU was handed.

In addition, the theoretical priority of project management needs in relation to M&E was not always put into practice by donor agencies. Donor information needs often overwhelm project M&E units. Ahmed and Bamberger (1989: 9) note 'the director of a large M&E unit ... [who] received a 50-page list of data requested by one donor. In another case a donor presented a long survey instrument to be completed within two weeks so that the data would be ready before the end of the supervision mission'. Coleman (1989e: 16) noted that 'those who compile MTE [Mid-Term Evaluation Reports] ... are often guilty — at one remove — of over-ambitiousness in terms of what they consider that the M&E system should provide'.

The Data Trap

The problem here has been neatly summarized by Chambers (1981: 97):

the extensive questionnaire survey with the 30 pages of questionnaire (multi-disciplinary, each discipline with its questions) which if asked are never coded, or if coded never punched, or if punched never processed, or if processed and printed out, never examined, or if examined, never analysed or written up, or if analysed and written up, never read, or if read, never understood or remembered, or if understood or remembered, never actually used to change action. Rural

504 *G. Coleman*

surveys must be one of the most inefficient industries in the world. Benchmark surveys are often criticised... and yet these huge operations persist, often in the name of the science of evaluation, pre-empting scarce national research resources, and generating mounds of data and papers which are likely to be an embarrassment to all until white ants or paper-shredders clean things up.

The evaluation literature — and the reminiscences of M&E practitioners — are replete with cases of massive survey exercises in which data were collected but not processed or analysed. Smith (1985: 108) referred to the problem that 'large quantities of unused (possibly unusable) data have been generated'; Clayton and Pétry (1983: 12) refer to M&E systems 'producing more data than are needed or can be processed'; Kinsey (1983: 156) refers to 'the generation, at some cost, of much unutilised data'; Slade (1983: 181) notes that the M&E system 'has generated a large amount of data, the processing and analysis of which has been subject to long delays'; Gow and Morss (1985: 181) referred to an evaluation survey for which 'The data analysis proved to be unmanageable... the bulk of it will probably never be used'; Coleman (1989d: 20) noted the two (separate) monitoring systems for agriculture in Mongolia, one of which generated 22,500 indicators annually for each agricultural enterprise. Most of the information from both systems was ignored for management purposes.

The individual case-studies provide a variety of reasons for these problems (often associated with poor computer performance and high staff turnover) but these 'downstream' problems are a consequence of the massive data collection efforts which are themselves often the consequence of unrealistic TOR for the M&E system. These unrealistic TOR may arise because 'it is often assumed that an M&E system (particularly once it is computerized) can collect information on almost any subject without significant cost or time implications' (Ahmed and Bamberger, 1989: 11).

Data Quality

The quality of data from these large data collection exercises is often poor. There are two components here concerning (a) the type of indicators used and (b) the implications of the methodology used to measure these indicators. These are clearly linked, but for the purposes of this paper they are treated separately.

Indicators

Agricultural and rural development projects have a limited implementation period (typically 5 years) during which time they are expected to achieve their production objectives. The measurement of this achievement is a major part of the M&E unit portfolio. In such projects crop yield (or crop production) and income increase targets are often specified and the TOR of the M&E unit will include their measurement.

Several problems arise in such measurements. Firstly, there is rarely an accurate baseline figure for these indicators (the 'before' of the classic before-and-after experiment) and baseline surveys undertaken during the project are often not completed, started years after the project start, or implemented so badly that the results are of little use (Chambers, 1981; Clayton, 1983; Coleman, 1989e; Gow and Morss, 1985). Secondly, those who set the TOR often appear to be unaware of just

Monitoring and Evaluation in Agricultural Projects 505

how difficult it is to measure these indicators. Income is notoriously difficult to measure amongst poor Third World farmers, especially if they have a vested interest in understatement (Coleman, 1988: 26). Casley and Kumar (1987: 133) note that 'in the case of small farmers it is difficult to define and measure income accurately; in fact it has rarely been done... if total farm or household incomes need to be measured, the difficulties become extreme'. Simpler proxy indicators of income (quality of housing materials, ownership of consumer durables, etc.), which are more in keeping with the limited resources available to most project M&E units, have not been popular with managers or planners, who prefer a 'harder' measure of progress.

The problems are even more intractable with crop yield or production measurement. Casley and Kumar (1987: 118) provide a table which indicates the number of years of high-quality production data which are required in order to detect a given trend (B) at given levels of accuracy and confidence. They show that in order to detect a rising trend of production of 4 per cent per time point, with an accuracy of 25 per cent either side, for 95 per cent confidence, we require 21 time points — equivalent to 21 years of data for annual cropping. Only one figure in the table is less than five time points. These figures lead Casley and Kumar to conclude (1987: 119) 'In stark language, the determination of yield or production trends in rain-fed smallholder farming areas may be impossible within the implementation period of most projects' (my emphasis). And yet these impossible measurements are still being included in the TOR of project M&E units, and managers are still demanding them.

Methodology

M&E staff are typically young and inexperienced, and the methodologies which are available to them are those from their undergraduate or postgraduate academic training. Emphasis here will all too often be on large-scale sample surveys and crop-cutting production/yield measurement. This neatly matches the evaluation bias of M&E TOR. Sample size considerations are often left out or dismissed in a few lines, resulting in 'bigger is better' samples. There is little on survey planning, which results in massive data collection exercises which take no account of the logistical requirements of the 'downstream' data processing and analysis stages, with results which have been examined previously. There is little emphasis on 'minimum' systems, rapid rural appraisal techniques, proxy indicators, cost-effective procedures. Little wonder then that with limited resources these M&E systems often fail to produce anything of use or value.

There is little in the academic training which emphasizes implementation or project beneficiary monitoring. The latter is particularly significant and has received some emphasis in recent M&E literature (Casley and Kumar, 1987). The traditional, evaluation-oriented, approach has been to track the changes resulting from project interventions via hard production indicators. Thus, for example, the examination of the effects of an agricultural extension project has typically involved efforts to measure the production benefits of each component of the extension package. Two implications follow from this approach.

Firstly, this approach typically involves large-scale sample surveys requiring the measurement of difficult input and production indicators. In addition, 'they attempted to show a causal relationship between extension services and yields that was analytically impossible to establish' (Murphy and Marchant, 1988: 6). Thus the

506 G. Coleman

approach was resource-intensive, intellectually demanding, and likely to produce disappointing results.

Secondly, this approach is remote from the target population. It asks what? but establishes why? only via econometric models derived from the survey data and a set of (often unstated) assumptions about the economic behaviour of the population. These models often 'failed to understand and take into account the motivations behind farmer's decision-making processes' (Murphy and Marchant, 1988: 6). In addition, these remote approaches have failed to contribute to, or gain from, 'participatory' M&E. This process involves gathering information from participants on their opinions concerning project inputs, activities and achievements. This provides a fast and cost-effective alternative to the production measurement approach noted earlier, and also provides direct and management-relevant feedback from project beneficiaries. It also provides the potential for much greater beneficiary involvement in the M&E process. Seeking beneficiary opinion implies a dialogue in which the project purpose is explained to beneficiaries in order to provide a context for their responses. Thus participatory M&E procedures are likely to be much more sharply focused and directly relevant to the intervention model being implemented.

Why do data collection activities take so long? Chambers (and many others) blame the (inappropriately) high professional standards of the collectors: 'Better, it is thought, to be long and legitimate than short and suspect' (1981: 99). The search for short-cut methods is hampered because 'the activities are not quite proper... [the practitioners] have a sense of responsibility to their professional training or more crudely they have been brainwashed by their professional conditioning' (1981: 98).

One element of this approach to methodology is an over-emphasis on sampling error at the expense of measurement error. Because sampling error can be measured in advance, while measurement error cannot, then samples are often designed to minimize the former while completely ignoring the latter.

Generally speaking an increase in the sample size for a survey reduces sampling error — but if at the same time it increases the workload on field and supervisory staff, or demands a large work force which is liable to be of lower quality overall, the quality of the data collection and processing work is likely to deteriorate and the non-sampling [measurement] errors to increase (Poate and Daplyn, 1990: 32).

Thus, for example, crop-cutting to measure yields may be subject to 'high levels of bias unless great efforts are made to supervise their execution' and 'Supervision of the enumeration quality of such... surveys is notoriously difficult' (Poate and Casley, 1985: 6-7). Indeed, Moore (1979: 4) noted that

The problem of controlling staff at a distance represents one of the major recurrent problems... in data collection the problem is especially acute because the falsification of work is so easy. The world is full of stories of farmer questionnaires filled up by the dozen in a quiet corner of an urban bar... a management consultant would probably identify staff control as *the* major issue for most data collection agencies.

Enumerator problems are by no means the only source of measurement error. In an examination of memory bias in labour data collection for small farmers in Nigeria,

Monitoring and Evaluation in Agricultural Projects 507

Coleman (1983) reported a mean level of over-reporting of labour inputs of 38 per cent.

5 THE SPECIFICATION OF INFORMATION NEEDS

One of the reasons why M&E systems tend to be 'maximum' rather than 'minimum' is that the supposed users — project managers — are rarely able to specify their information needs in advance. Thus Gow and Morss (1985: 181) note that 'key project personnel often find it difficult to specify in advance what information they need for planning, monitoring and evaluating project activities', and Casley and Kumar (1987: 22) note that

more often than not the users themselves have not decided what they really need. The experience of many projects shows that managers describe topics or issues only in general. One often hears the remark 'We want all relevant information about project implementation'. Designers of the information system should therefore try to help managers identify specific items of information.

One result of the above is that it is the M&E officers themselves who attempt to specify the information needs, since 'Management is not involved in the design of the M&E system' (Deboeck and Kinsey, 1980: 16). Lacking guidance, these M&E implementors (young, inexperienced) often fall back on their inappropriate training and mount large surveys. To the extent that these surveys matched the designs of the planners, the ambitions of the M&E TOR, and covered the needs of the donors, they were generally well received, at least until they failed to deliver.

6 MANAGEMENT USE OF INFORMATION

Chambers (1981: 95) noted that 'Decision-makers need information that is relevant, timely, accurate and usable. In rural development, a great deal of the information that is generated is, in various combinations, irrelevant, late, wrong and/or unusable anyway.' Deboeck and Kinsey (1980) show how some of these issues contribute to a substantial rift between management and the suppliers of information (see also French and Walter (1984) for a detailed examination of the way in which evaluation findings failed to affect project implementation).

This problem arises in part from features noted earlier: over-ambitious M&E designs frequently fail to produce usable results on time, often because of data processing delays; the focus on evaluation does not serve project management needs; the methodology used is frequently inappropriate to the resources available, resulting in poor data quality; managers demand the measurement of difficult indicators; managers fail to specify their information needs. In addition,

Management is presented with half-digested data rather than usefully interpreted information... the information is inconclusive or is not presented in an intelligible form... management lacks confidence in the information... the information does not agree with preconceived views or hypotheses (Deboeck and Kinsey, 1980: 17).

508 G. Coleman

All of these result in a poor service to managers, who are thus more likely to dismiss the M&E system as irrelevant to their needs, or view it as a servant of the donors.

These *technical* issues are serious enough, but perhaps not as significant as political issues. Project management frequently perceives the M&E system as a spy or a policeman, forced on the project by outsiders and designed to report on the performance of management to central ministries and donors (Ahmed and Bamberger, 1989; Casley and Kumar, 1987; Chambers and Belshaw, 1973; Clayton, 1983; French and Walter, 1984; Gow and Morss, 1985). Almost all M&E practitioners have examples of findings which are ignored, camouflaged or actively suppressed by management. Almost all major M&E texts and reviews recognize this issue, and recommend as treatment the greater integration of M&E with project management, and greater efforts to assist management in specifying its own information needs. This may help, but 'it would be naïve to ignore that an effective information system — particularly its monitoring and evaluation functions — can be seen as a threat by project management' (Gow and Morss, 1985: 180). And not only project management: 'Donors are hesitant to act on project information, particularly if it calls for midterm corrections in project activities or for their termination' (Gow and Morss, 1985: 175-176). Technically superior M&E may still fail if this issue cannot be effectively addressed.

7 CONCLUSION

M&E systems in agricultural and rural development projects have frequently not performed the task which was set for them. However, the careless gloom with which their efforts are often dismissed is at least partly a reflection of the too-high expectations at their inception. M&E as the single necessary 'fix' for problematic project management was always unrealistic. Projects fail for many more reasons than the lack of M&E.

These too-high expectations have been carried over to the TOR of M&E systems. Their ambition reflects a failure to specify management information needs, a misplaced faith in the ability of classroom methodologies to work in the field, an unrelenting demand for the measurement of the difficult or impossible, and a naïve belief in the ability of young and inexperienced M&E officers to carry out a wide set of large-scale, management-intensive data collection activities with few resources. Again, this was always unrealistic.

It is perhaps ironic that systems supposedly designed to learn from experience have been so slow to learn from their own. However, in recent years there have been encouraging signs. The interest in Rapid Rural Appraisal (Chambers, 1981) partly stemmed from the experience of academics and field-workers who — outside formal M&E systems — had developed cheap and efficient monitoring and evaluation methods. It also arose from the experience of early participants in formal M&E systems, who watched their spectacular information machines grind to a halt, choking with their own data. It was a salutary experience.

Methodology manuals now offer much better guidelines in 'real-world' fieldwork practice, including the use of proxy indicators, and with much greater awareness of resource constraints (Casley and Kumar, 1988; Poate and Casley, 1985; Poate and Dapiyn, 1990). M&E designs now eschew the 'traditional' large-scale, cover-everything socioeconomic samples (Coleman, 1989a) and concentrate instead on agreeing

Monitoring and Evaluation in Agricultural Projects 509

information needs with users (Tame, 1989) and specifying 'minimum' systems with respect to the number and type of indicators, and the data collection methodology (Coleman, 1989e). A notable example of the latter has been the recent work on the monitoring and evaluation of agricultural extension projects (Coleman 1989a,b,c; Marchant, 1983, 1986; Murphy and Marchant, 1988).

There is still a great deal of learning to be done in this area. New recruits are entering the field and will inevitably re-invent the old wheel (inevitably because their training provides for no other approach). However, as the old, somewhat battered, hands move up to project management they will, hopefully, be better prepared to play a more active role in specifying their information needs, and in making effective use of the information provided.

REFERENCES

- Ahmed, V. and Bamberger, M. (1989). *Monitoring and Evaluating Development Projects: The South Asian Experience*. Washington: EDI Seminar Series, World Bank.
- APMEPU (1982). *Funtua Agricultural Development Project Completion Report*. Kaduna, Nigeria.
- Baldwin, K. D. S. (1957). *The Niger Agricultural Project*. Oxford: Basil Blackwell.
- Casley, D. J. and Kumar, K. (1987). *Project Monitoring and Evaluation in Agriculture*. Baltimore: Johns Hopkins University Press for the World Bank.
- Casley, D. J. and Kumar, K. (1988). *The Collection, Analysis and Use of Monitoring and Evaluation Data*. Baltimore: Johns Hopkins University Press for the World Bank.
- Chambers, R. (1981). 'Rapid rural appraisal: rationale and repertoire', *Public Administration and Development*, 1, pp. 95-106.
- Chambers, R. and Beishaw, D. G. R. (1973). 'Managing rural development: lessons and methods from eastern Africa', IDS Discussion Paper No. 15, Institute of Development Studies, University of Sussex.
- Clayton, E. (1983). 'Role, characteristics and operational features of agricultural monitoring systems', In Clayton, E. and Pétry, F. (eds), *Monitoring Systems for Agriculture and Rural Development Projects*, vol. 1, pp. 1-19. Rome: FAO Economic and Social Development Paper 12 REV 1. Two volumes.
- Clayton, E. and Pétry, F. (eds) (1983). *Monitoring Systems for Agriculture and Rural Development Projects*. Rome: FAO Economic and Social Development Paper 12 REV 1. Two volumes.
- Coleman, G. (1983). 'The analysis of memory bias in agricultural labour data collection: a case study of small farmers in Nigeria', *Journal of Agricultural Economics*, 34, pp. 79-86.
- Coleman, G. (1988). *Ethiopia-IFAD Agricultural Credit Project: Mid-Term Evaluation Report*. Rome: Overseas Development Group for IFAD.
- Coleman, G. (1989a). *Fayoum Agricultural Development Project: Monitoring and Evaluation Proposals*. Rome: Overseas Development Group for IFAD.
- Coleman, G. (1989b). *Minia Agricultural Development Project: Monitoring and Evaluation Proposals*. Rome: Overseas Development Group for IFAD.
- Coleman, G. (1989c). *West Beheira Settlement Project: Monitoring and Evaluation Proposals*. Rome: Overseas Development Group for IFAD.
- Coleman, G. (1989d). *An Agricultural Information System for Mongolia*. Rome: FAO.
- Coleman, G. (1989e). *Analysis of the Performance of the Monitoring and Evaluation Systems: Arrangements in IFAD-Supported Projects*. Rome: Overseas Development Group for IFAD.

510 G. Coleman

- Coleman, G. (1990). 'Problems in project-level monitoring and evaluation: evidence from one major agency', *Journal of Agricultural Economics*, 41, pp. 149-161.
- Deboeck, G. and Kinsey, B. H. (1980). 'Managing information for rural development: lessons from eastern Africa'. Washington: World Bank Staff Working Paper No. 379.
- Forrest, T. (1981). 'Agricultural policies in Nigeria 1900-1978'. In Heyer J., Roberts, P. and Williams, G. (eds), *Rural Development in Tropical Africa*. London: MacMillan, pp. 222-258.
- French, W. and Walter, M. (eds) (1984). *What Worth Evaluation?* Boroko, Papua New Guinea: Monograph 24, Institute of Applied Social and Economic Research.
- Gow, D. and Morss, E. (1985). 'Ineffective information systems'. In Morss, E. and Gow, D. (eds), *Implementing Rural Development Projects: Lessons from AID and World Bank Experiences*. Boulder, Colorado: Westview Press, pp. 175-197.
- Hesling, L. (1984). 'A note on impact monitoring of agricultural development projects', *Journal of Agricultural Economics*, 35, pp. 279-281.
- Hunter, G., Bunting, A. H. and Bottrall, A. (eds) (1976). *Policy and Practice in Rural Development*. London: Croom Helm.
- Kinsey, B. H. (1983). 'Monitoring large-scale agricultural development projects: Lilongwe, Malawi. In Clayton, E. and Pétry, F. (eds), *Monitoring Systems for Agriculture and Rural Development Projects*, vol. 2, 155-180. Rome: FAO Economic and Social Development Paper 12 REV 1. Two volumes.
- Lele, U. (1975). *The Design of Rural Development*. Baltimore: Johns Hopkins University Press for the World Bank.
- Marchant, T. J. (1983). 'The Kenya National Extension Program: monitoring and evaluation system'. Unpublished manuscript.
- Marchant, T. J. (1986). *Egypt-Minia Agricultural Development Project: Monitoring and Evaluation Programme*. Rome: Report No. 0097-EG, IFAD.
- Moore, M. (1979). 'Denounce the gang of statisticians; struggle against the sample line: unite the researching masses against professional hegemony', Rapid Rural Appraisal Conference, Institute of Development Studies, University of Sussex.
- Murphy, J. and Marchant, T. J. (1988). 'Monitoring and evaluation in extension agencies'. Washington: World Bank Technical Paper No. 79, Monitoring and Evaluation Series.
- Poate, C. D. and Casley, D. J. (1985). *Estimating Crop Production in Development Projects: Methods and Their Limitations*. Washington: World Bank.
- Poate, C. D. and Daplyn, P. F. (1990). 'Data for agrarian development'. Unpublished draft.
- Slade, R. (1983). 'Monitoring agricultural development projects: the Nigerian experience'. In Clayton, E. and Pétry, F. (eds), *Monitoring Systems for Agriculture and Rural Development Projects*, vol. 2, pp. 181-203. Rome: FAO Economic and Social Development Paper 12 REV 1. Two volumes.
- Smith, P. J. (1985). 'Monitoring and evaluation of agricultural development projects: definitions and methodology', *Agricultural Administration*, 18, pp. 107-120.
- Tame, J. A. G. (1989). *Management Information System: Final Report*. Bamenda, Cameroon: North West Development Authority (MIDENO).
- van de Laar, A. (1980). *The World Bank and the Poor*. London: Institute of Social Studies, Series on the Development of Societies, vol. VI, Martinus Nijhoff.
- World Bank (1974). *Appraisal of Funtua Agricultural Development Project*. Washington: Report No. 345a UNI.
- World Bank (1988). *Rural Development: World Bank Experience 1965-86*. Washington: Operations Evaluation Department.
- Yudelman, M. (1976). 'The World Bank and rural development'. In Hunter, G., Bunting, A. H. and Bottrall, A. (eds), *Policy and Practice in Rural Development*. London: Croom Helm, pp. 21-29.

A WORLD BANK SPONSORED STUDY
ON

INDIA : PERFORMANCE OF STATE MONITORING
AND EVALUATION UNITS UNDER THE TRAINING
AND VISIT SYSTEM OF AGRICULTURAL EXTENSION

SUMMARY

BY

JAI KRISHNA

S.K. RAHEJA

CENTRE FOR AGRICULTURAL AND RURAL DEVELOPMENT STUDIES
NEW DELHI

APRIL 1994

INDIA : PERFORMANCE OF STATE MONITORING AND EVALUATION UNITS
UNDER TRAINING AND VISIT SYSTEM OF AGRICULTURAL EXTENSION

SUMMARY

1. Background

Reorganising the agricultural extension approach on the lines of the T&V system in India was easily the most elaborate exercise of its type undertaken anywhere in the world. Between 1977 and 1982, 74 million farm holdings representing 76.2% of the total number of holdings were brought under the coverage of the T&V system. Another 22 million farm holdings were brought under the reach of the T&V extension system between 1985 and 1987. To monitor the reach of the re-organised extension system covering such a large number of farm holdings alone was a huge task. Monitoring the adoption of impact points/main recommendations and their effects made the task truly gigantic.

Thanks to the support provided by IDA/WB, both financial and technical, the response of the states and the Central (GOI) Directorate of Extension was equally impressive. As for the setting up of the Monitoring and Evaluation Units (MEU) is concerned, out of the 13 states covered under the T&V programme in the first phase (1977-1982), MEUs were set up very quickly (within one year) in 11 states. Following the preparation of the Instructions Manual (1) in 1981 and the training workshops organised by the WB/DOE, the stage was set for systematically monitoring the reach of the re-organised extension system. Thus the installation phase of MEUs was handled with a great deal of success.

Then followed the active phase of about 5 to 7 years (1982-89) of MEUs' work and their usefulness. The introduction of a highly systematic and sharply focussed approach of agricultural extension generated a great deal of interest and enthusiasm among the extension workers and their senior managers in the states covered under the T&V programme. The findings of M&E surveys and studies provided a ready reckoner to gauge the reach of the re-organised extension system and its initial effects (adoption of impact points). Given the users' interest and appreciation, the response of managers and staff of MEUs was also prompt and positive. Regular monitoring and, in several cases, M&E surveys were conducted and the findings were put out within acceptable time limits. In several states, an impressive number of special studies were also conducted during this period. Extension managers too responded by initiating corrective action.

However, after about 10 years of experience, both with the T&V system and the highly structured monitoring and M&E effort, clear signs of 'fatigue' and 'inertia' became visible. Visit rates of VEWs started to stagnate, diffusion of technology from the CFs to DFs remained weak, monthly workshops and fortnightly trainings became routine affairs and research-extension linkage lost a lot of its shine. This phase also saw a sharp decline in the interest of the senior extension managers in 'routine' and, in the words of several top extension managers, 'stereo-typed' indicators of the reach of extension and adoption of impact points. Interaction between the MEUs and the users of their findings came under severe strain due to the somewhat adverse trends seen in the working of the T&V system. In some states the

Consultants were even told that the MEUs were being pressurised to suitably adjust their findings to reflect the achievements of the extension service in a positive light. While the Consultants would not like to confirm any manipulation of MEUs' findings, in some cases, there are unmistakable indications to suggest that the reported reach of the T&V extension system does not reflect the ground realities.

The past couple of years (1992-1993) may be termed as a phase of dilution and decline in the activities and effectiveness of MEUs. Following the phasing out of the IDA/WB support to all the 13 Phase I (1977-82) states in 1993, and in some cases even earlier, a number of MEUs have experienced neglect and isolation. Staff vacancies are not being filled, MEUs' transport facilities have either become unusable or have been made part of the departmental pool of vehicles, funds for operational expenses (travelling, maintenance of equipment, stationery, etc.) are being severely curtailed and in several cases, MEUs' staff have been assigned for other duties. But for the promise of further IDA/WB support under the follow-up composite Agricultural Development Projects, MEUs in at least four out of the eight states covered under this study may have become totally non-functional by now.

2. Assessment of MEU's Performance and related characteristics:

The above adverse trends are clearly reflected in the performance rating of the selected MEUs (Chapter 4 of the main report). Table S.1 summarises the Consultants' rating of MEU's performance and the related characteristics for the selected MEUs. It also gives the trend in performance in relation to 1990(+).

Out of the seven MEUs whose performance has been assessed by the Consultants, four present an unmistakable picture of decline. In the case of the Madhya Pradesh MEU, no change in effect means that the stage of complete drift and ineffectiveness visible in 1990 has continued. The only exceptions are Karnataka and Tamilnadu. In Karnataka, the good standard of performance has been maintained. There are firm indications that the Karnataka MEU will soon be placed under the Director of Agriculture. In Tamilnadu, while no change is recorded in the already good performance of the MEU as observed in 1990 (2), there has been a definite improvement in the organisational status of the Unit.

The overall performance of MEUs and the aggregated score of the main characteristics affecting their performance, as presented in Table S.1, are evidently well correlated. A good overall status of the characteristics identified in this report is likely to lead to good overall performance also. The MEUs that have performed well (Haryana, Karnataka and Tamilnadu) have also been the beneficiaries of good overall managerial and user support. On the other hand, poor support for the characteristics (identified in Chapter 3 of the main report) has also resulted in poor performance by MEUs, e.g., in M.P., Maharashtra and Orissa.

(+) In 1990 also, the writers of this report were involved in the CARDS/DOE study on 'Functioning of State MEUs'. Although no formal scoring/rating was attempted at that time, the Consultants did develop a fairly good idea of the relative performance of the MEUs. All the MEUs covered in the present study were also covered in the 1990 study.

Table S.1: Assessment of MEUS' performance and related characteristics

State	IDA/WB support to MEU period	Performance rating (score)	Rating of related Characteristics (score)	Recent trend in performance 1990-93	Remarks
1	2	3	4	5	6
1.Haryana	1979-1993	Good (64)	Good (70)	Declining	Senior MEU posts have not been filled. Effectiveness of MEU also appears on the decline.
2.Karnataka	1980-1993	Good (65)	Good (64)	Constant	MEU is likely to be placed under Director of Agriculture. Recent effort of MEU to modify TBV approach to TVED needs to be carefully evaluated. Special studies have been recommended.
3.Madhya Pradesh	1978-1993	Very poor (20)	Poor (40)	No change	In 1990 also MEU was in a poor shape. Since then there has been no improvement. The Unit remains ineffective.
4.Maharashtra	1981-1987	Average (a the margin -45)	Average (55)	Declining	Reorientation of extension focus to target based special thrust programmes and lack of reflection of this change in MEU's surveys has made the Unit ineffective.
5.Orissa	1986-1993	Very poor (28)	Poor (44)	Declining	It took 9 years after inception of TBV (1977) to set up the MEU. The Unit was about to be disbanded in 1993. The expectation of follow-up WB assistance is keeping it alive.
6.Uttar Pradesh	1987 onwards	Very poor (23)	Average (50)	Declining	In 1990 the Unit was in its infancy. In spite of its advantageous placement, the Unit has been ineffective. The staff and funds of MEU have not been properly deployed.
7.Tamil Nadu	1981-1987	Very Good (75)	Very Good (77)	Improving	In spite of a number of risk factors, the MEU in Tamil Nadu has performed very well. The Unit has now been expanded and redesignated as Policy, Planning and Management Wing of the Directorate of Agriculture. PPMW is one of the front line units of the Directorate.
8. Punjab	1987 onwards				Since MEU has not been set-up in Punjab, its performance has not been assessed. However, all indications point to user indifference.

It is also of interest to examine how the overall performance of MEUs correlates with the main characteristics affecting performance. Accordingly, an attempt has been made to study the correlation between performance and the characteristics disaggregated into four main groups, namely: a) organisational aspects b) personnel matters c) support facilities and d) MEU-user interactions. Table S.2 presents the relevant information.

Table S.2: MEU's performance and related groups of characteristics

State	Performance rating (*)	Rating of characteristics group			
		Organisa- tional aspects	Personnel aspects	Support facilities	MEU-User interaction
1.Haryana	B	A+	B	C	C
2.Karnataka	B	B	C	B	B
3.M.P.	D-	C	D-	D	D
4.Maharashtra	C	B	C	B	D
5.Orissa	D-	B	D	D	D
6.U.P.	D-	B	D-	B	D
7.Tamilnadu	A	A+	B	A	A

(*) A+ = Excellent; A = Very good; B = Good; C = Average; D = Poor; D- = Very poor

It will be seen that the different characteristics groups are not equally strongly correlated with performance. In fact, the organisational aspects group, which includes hierarchical placement of MEU, rank of head of MEU and internal structure of the Unit, has a weak correlation with performance. Some states that fare well on this group's score (Maharashtra, Orissa and U.P.) have not performed well. This would indicate that this group of characteristics does not have a strong explanatory value in so far as an MEU's performance is concerned. In this context it is worth recalling the finding of the earlier CARDS study (2). The study had concluded that "There appears to be no direct relationship, as assessed by the Study Team, between the organisational structure of an MEU and its performance". However, the ratings presented in the Table 5.2 indicate that a good organisational and hierarchical pattern provides a good environment for an MEU to succeed, as is the case in Tamil Nadu, Haryana, and Karnataka. By all accounts the personality of head of the Unit, a trait that cannot be easily quantified, has a significant influence on the performance of an MEU.

The personnel matters group - staff strength (actual to sanctioned), staff skills, stability of tenure and recruitment and promotion policies, has a very strong correlation with performance. MEUs that have performed poorly have invariably been the victims of poor personnel policies and management. Tamil Nadu is a special case where the only group to receive a 'B' rating is the 'personnel matters'. This issue has been examined in

section 4.7 of the main report.

Support facilities group does not seem to have a strong correlation with performance. However, thanks to IDA/WB support, in very few cases have the MEUs been seriously handicapped for want of funds, transport and even data processing facilities.

MEUs' interaction with the users comes out as a very strongly correlated group with performance. This is how things really are. Wherever top administrators and senior extension managers (Secretary, Director and Additional Director) have taken, or have been motivated to take, active interest in MEU's work, it has acted as a tonic for the whole system. Wherever they have been indifferent, MEUs have faced neglect and isolation.

3. Some important issues and lessons

A number of important issues emerge from this study. First, there is the riddle of why some MEUs were facilitated to perform well while the others were clearly put to a disadvantage. It is well known that the T&V programme had a highly standardised format of focus and approach throughout the country. Consequently, the focus and approach of M&E work also followed a standardised format. Moreover, the provision made for MEUs (staff, transport, operating expenses, etc.) under the IDA/WB projects were also similar for all the states. What, then, explains the phenomenon of MEUs in some states performing well and in others faring poorly?

The Consultants have no ready answers. The angle of a Civil Service (IAS) Director vs a technocrat Director has been explored. Among the well performing states, Haryana and Tamilnadu have IAS Directors, whereas Karnataka has a technocrat. The states that have performed poorly - M.P., Orissa and U.P., all have technocrat Directors. On the other hand, the Maharashtra MEU has gone downhill very fast under IAS Directors. Thus no firm pattern emerges. It may, however, be mentioned that an M&E system which originates outside the Directorate but is implemented through it has a poor chance of success under a technocrat Director. Due to his/her intimate knowledge of the state's agriculture, a technocrat is not likely to be impressed with routine monitoring indicators which remain unchanged for long years. In many cases, these officers have their own non-formal sources of getting feed back. Moreover, the long years a technocrat Director would have put in the Directorate before getting to the top post make him/her susceptible to certain biases and likes and dislikes. The only way to ensure success in such cases is to involve the technocrats right from the planning stage of an MEU. A serious effort should be made from the beginning to sensitize the technocrats of the potential of an efficient M&E system as a powerful management aid. A series of orientation seminars should be planned for the top managers at the beginning and at periodic intervals. Finally, the Directors must be closely involved in specifying the scope and focus of work of MEUs. As long as M&E work is seen as a World Bank or GOI requirement, the technocrats may be passive. To ensure success, the M&E work must be planned and executed as part of the feed back requirement as perceived by the Director and his senior officers.

A number of other lessons emerge from the Indian experience. It is strongly felt that there should be clarity about the organisational objectives of MEU. If an MEU is set up solely for the specific purpose of providing feed back during various phases of a project/programme, the personnel and organisational policies must be designed accordingly and the M&E Unit should be disbanded at the end of the project. On the other hand, if an MEU has to move from a 'specific' to a 'general' M&E Unit for the Directorate, the personnel policies and organisational structure must clearly reflect this objective at all stages of the programme. In India, most of the MEUs are now in a state of animated suspension. They are neither part of the Directorate's hierarchy, nor are they out of it. The only exceptions are Haryana and Tamilnadu, where the role of MEUs has been fully internalised. In Karnataka, the MEU is on the verge of being fully integrated in the Directorate of Agriculture.

Another important lesson relates to clarity of M&E concepts. Under the T&V programme, M&E concepts have been loosely defined and used. While the dividing line between monitoring and evaluation is admittedly thin, it is necessary to clearly separate the two concepts. The World Bank has provided very valuable guidelines for this purpose (3). In this context, it has been suggested to the Consultants that monitoring should be an internal management function of the principal users, i.e., the Director of Agriculture and his senior colleagues, whereas the responsibility of evaluation, including concurrent evaluation, should be carved out to either a separate cell under Secretary (Agriculture) or to the Planning Ministry of state governments. Combining both the functions under a single unit has not worked well and it is unlikely that it can work well.

The issue of both monitoring and evaluation being dynamic concepts has been discussed in considerable detail in Chapter 3 of the main report. As has been pointed out the focus and emphasis in agricultural extension and hence in M&E work must change as agriculture absorbs progressively higher order of technology and gets increasingly commercialised. The routine, ubiquitous monitoring indicators of VEW-farmer contacts, rate of adoption of oversimplified recommendations, etc lose relevance over time. The feedback required by the principal users shifts to the higher plane of diagnostic study of constraints in optimal resource use, efficiency and cost-effectiveness of improved technology, etc. The inability of MEUs in India to appreciate the dynamic aspects of the M & E focus has most certainly resulted in 'user-indifference'. In most of the states the Directors of Agriculture readily expressed their disappointment at being presented the same routinised M & E information year after year. The relevance of maintaining the distinction between the CFs and OFs has also been openly questioned; yet it has not been dispensed with in M&E work. In the case of an agriculturally progressive state the low visit rate of VEWs was dismissed by the Director as of no consequence since even the raw effects indicators—sale of certified seeds, fertilisers, etc, pointed to a healthy uptake of technology by the farmers. Similarly, in another state farmers receiving fewer VEW visits were recording higher adoption/yield rates; yet, no inferences were drawn by MEU.

The fact of the M & M&E reports getting delayed for 18 months to over two years in most of the states covered in this

study without inviting loud protests from the main users is another pointer of low user interest in routine M & E indicators. The Indian experience clearly indicates that the MEU's managers have to remain alert to the changing needs of the users. Involvement of the State Agricultural Universities and outside Consultants in periodically reviewing the M and E needs of the users and in designing suitable methodology and approach should prove useful.

Matters relating to MEU personnel, including their basic qualifications and experience, their skills-mix, career prospects, in-service training, stability of cadre and tenure, etc., need to be addressed with utmost seriousness. The tendency to use MEU posts either as promotion avenues or as a dumping ground for unwanted officers has to be eschewed. A competent multi-disciplinary team of officers goes a long way in ensuring success of an MEU.

A related aspect is the weakness of the analytical techniques used by even the well performing MEUs. Conclusions based on comparisons of simple averages and ratios may be misleading. The senior officers of MEUs should have a good grounding in simple techniques of statistical analysis, such as of significance of difference of means, analysis of variance, regression analysis, etc. Similarly, a good orientation in simple tools of economics analysis, farm planning and budgeting is also necessary.

The need to conduct sample surveys covering the whole state during every crop season should also be re-assessed. Less costly techniques like rapid appraisal assessment, rotating surveys covering a given proportion of districts (e.g. 1/3 rd) every year, should be seriously explored. The State Agricultural Universities can play a valuable role in providing short-term training and consultation to MEUs on the above aspects.

Another important issue relates to the administrative misuse of the M&E findings. Instances have come to the notice of the Consultants to the effect that indicators like poor VEW visit rate, low adoption rates, etc., have been used as a basis for adverse remarks in the annual confidential reports of the district officers. In some cases, even the Directors are reported to have suppressed some unpalatable findings due to fear of audit queries. The problem gets compounded with the highly hierarchy oriented administrative structure in India which relies on outdated methods of appraisal of staff performance. In this setting any unit purporting to report of the achievements of another unit is bound to invite suspicion, if not outright hostility.

While the instances of mis-use of M&E findings are very few and far between, they represent a disturbing development. Monitoring, as indicated earlier, is essentially an aid to management. It should never be mixed up with performance appraisal of individual staff members. All concerned, particularly the top bosses, auditors and even the politicians must realise that M&E findings cannot be used to penalise officers for long. In such an event, the M&E data will begin to sing a different tune.

R E F E R E N C E S

1. Slade R.H. and G. Feder, The Monitoring and Evaluation of Training and Visit Extension in India: A Manual of Instructions, The World Bank, Washington D.C., October 1981.
2. Centre for Agricultural and Rural Development Studies (CARDS), Functioning of State Monitoring and Evaluation Units - A Special Study, Directorate of Extension, Ministry of Agriculture, New Delhi, 1991.
3. Casley D.J. and Krishan Kumar, Monitoring and Evaluation in Agriculture, The World Bank, Washington D.C., 1987.