Evaluation in World Bank Education Projects: Lessons from Three Case Studies

Barbara Searle, Editor
August 1985

The views presented here are those of the author, and they should not be interpreted as reflecting those of the World Bank.
The World Bank does not accept responsibility for the views expressed herein which are those of the author and should not be attributed to the World Bank or to its affiliated organizations. The findings, interpretations, and conclusions are the results of research or analysis supported by the Bank; they do not necessarily represent official policy of the Bank. The designations employed, the presentation of material, and any maps used in this document are solely for the convenience of the reader and do not imply the expression of any opinion whatsoever on the part of the World Bank or its affiliates concerning the legal status of any country, territory, city, area, or of its authorities, or concerning the delimitation of its boundaries, or national affiliation.
The four papers in this volume review and discuss the evaluation components in three World Bank financed education projects and draw lessons about the use of evaluation in projects and development activities in general. The experience reviewed suggests that, in development projects: (1) evaluation resources are often more effectively used to gather information to improve project performance rather than to assess project impact; (2) the type of evaluation strategy adopted depends crucially on the degree to which a stable intervention is in place, as contrasted with new activities undergoing trial, and whether or not the inputs being delivered are themselves sufficient to bring about the expected outcomes; (3) conceptual modeling of how an intervention is expected to produce the intended outcomes is important for effective implementation and absolutely essential for conducting an evaluation; and (4) the institutional setting where an evaluation unit is housed is a crucial factor in determining whether and in what ways the information gathered will be used.
ABOUT THE AUTHORS

JOHN MIDDLETON's academic and professional work has centered on the planning, management and evaluation of communication systems, with special focus on educational technology. Author or editor of five books and a number of articles in these areas, Middleton was with the East-West Center and the Academy for Educational Development before becoming the Director of the School for International Training in 1983. He has consulted on education and communication planning and evaluation in the USA, Philippines, Indonesia, Zimbabwe, Egypt and Oman. From 1980-82 he was Planning Advisor for communication technology to the Indonesia Ministry of Education.

PAUL A. SCHWARZ is President of the American Institutes for Research. He has served as the Director of the Institutes' International Division from 1960 to 1970, supervising projects in 19 developing countries. As part of this work, he served a four-year tour in West Africa and a two-year tour in Thailand.

RICHARD SACK is an independent consultant in the areas of educational planning, policy analysis, and the management of policy formulation and implementation. He received his Ph.D. (1972) in International Development Education from Stanford University. He has taught in universities in the United States, Canada, France and Zaire as well as in Algerian secondary schools. He has conducted research in Tunisia and Zaire and has worked in a number of countries, mainly in Africa and the Caribbean. At present he lives in Paris and consults mainly with international organizations.

JOHN RAVEN's main interest is in values, attitudes, and institutional structures associated with economic and social development. He has been involved in socio-economic research to clarify the wider goals of general education, in evaluating the extent to which these goals are achieved, in studying the reasons why they are not achieved more effectively, and in identifying ways in which the necessary competencies and understandings can be fostered by parents, teachers, and managers. He believes Education, Values and Society (1977) and Competence in Modern Society (1984) to be his most significant publications.
# TABLE OF CONTENTS

Overview.............................................................................................................1

Section I: The Philippine Radio Education Project:
A Case Study of Decision, Design and Interactions .................. 7
by John Middleton

Preface.............................................................................................................9
A Note on Names..........................................................................................10
Introduction.................................................................................................11
Project Initiation..........................................................................................14
Formative Year..............................................................................................17
Pilot Year and Beyond.................................................................................33
Endgame.........................................................................................................50
Patterns and Lessons....................................................................................55
Notes..............................................................................................................63
Appendix.........................................................................................................67

Section II: Evaluation of the Pakistan Primary Education Project:
A Methodological Case Study by Paul A. Schwarz.........................73

Preface.............................................................................................................75
The Methodological Context.................................................................77
The Primary Education Project.............................................................80
The Research and Evaluation Component............................................84
The Evaluation Component Rethought.................................................85
Conduct of the Reconnaissance.............................................................88
Implications for the Project in Pakistan.................................................101
Implications for Other Projects...............................................................102

Section III: Lessons From An Evaluation/Monitoring System
for the Haitian Education Reform by Richard Sack.........................107

Introduction.................................................................................................109
A Summary Review of the Reform..........................................................111
The Evaluation/Monitoring System.........................................................116
The Impact of the Evaluation/Monitoring System on Policy.................128
Conclusion....................................................................................................132
Lessons..........................................................................................................134
Annex: A Practical Follow-up in the Realm of Design.........................135

Section IV: Some Questions and Lessons for the Bank
from the Three Case Studies by John Raven........................................139

Introduction.................................................................................................141
The Role of the Evaluator........................................................................143
The Case for the "Illuminative" Approach...............................................147
Barriers to Effective Illumination............................................................150
Arrangements for Policy Research..........................................................156
Conclusion....................................................................................................166
References...................................................................................................170
Since the early 1970s the World Bank has recognized the importance of evaluation in development projects; during the past decade both the number of evaluation activities included in education projects and the level of financing has been increasing. As part of an effort to learn from this project experience, the Education and Training Department commissioned case studies of evaluation activities in three projects financed by the Bank. This volume presents the cases and a "think piece" that draws on the findings of the cases to assess the role of evaluation in development in general, and Bank-financed projects in particular.

Evaluation of social programs received its greatest impetus from the Great Society programs in the United States. During those early years of substantial funding and high hopes, evaluation was looked to primarily to assess project impact. The prime question was the "summative" one: Did a project attain its objectives and could the outcome be attributed unequivocally to project activities? Answering this apparently simple question about projects being implemented in field settings, where nothing works as expected, and if something can go wrong, it will, has proven unexpectedly difficult. These difficulties have stimulated two quite different responses. On the one hand, the past several decades have seen an explosion of new research techniques and sophisticated methodologies designed to deal with the difficulties of doing rigorous scientific work in such settings. The other response has been to step back and inquire: Is
the "summative" question about project impact the question we should be asking? Or, alternatively, should evaluation resources be used to address questions about project implementation, or to make predictions about potential outcomes, or to put in place monitoring systems that will keep project staff in touch with field activities. The papers presented here explore this issue from a variety of perspectives, drawing on experiences in quite different settings.

The Philippine Radio Education Project was designed very much in the style of the Great Society programs. It was explicitly a pilot project and the level of funding for evaluation was very high -- close to half of project costs. The evaluation design was methodologically sophisticated and it was carried out with remarkable fidelity by highly competent Philippine researchers. In his study, John Middleton carefully traces the life history of the evaluation, its degree of ownership by crucial actors in the Philippines and the Bank, and the role the evaluation findings played in subsequent decisions about investment in primary education in the Philippines. He documents both the strengths and weaknesses of addressing the summative question.

The Pakistan Primary Education Project, while less explicitly a pilot project, nevertheless envisioned the same role for evaluation: that evaluation findings would demonstrate which inputs were effective in raising the quality of primary education and would thus lead directly to the design of an effective follow-on project. For a variety of reasons, the evaluation work did not progress well, and the Bank asked Paul Schwarz to supervise and provide technical assistance to that component of the project. Schwarz's paper describes both why and how he helped the Pakistanis restructure the evaluation. In his analysis of the role of
evaluation, Schwarz characterizes projects along two dimensions: (i) whether the intervention (that is, the inputs and how they are used) is in a stable form or still being modified, and (ii) whether the inputs that the project is providing are sufficient or only necessary conditions for producing the benefits sought. These distinctions help Schwarz address the dilemma about where evaluation resources can most effectively be used.

The aspirations for evaluation are much broader in the Haiti educational reform, which is being financed in part by a series of credits to the Government of Haiti. Provision was made for monitoring the productivity of key educational inputs, for evaluating the effectiveness of several key elements of the reform, and for assessing the reaction of key audiences, namely teachers and private schools, to new policies and procedures. In part because these activities were to be carried out by different organizations, and in part because of the comprehensive nature of the reform, Richard Sack's study focuses on the political and institutional aspects of evaluation. He documents the need for, and the difficulties inherent in, addressing the broad range of questions noted above, that go far beyond the issues of project impact. His study draws attention to the managerial complexity of establishing evaluation units whose work will be accepted by technical staff and also serve the needs of policy makers and the public at large. While the evaluation work in Haiti is technically less sophisticated than that undertaken in the Philippines or Pakistan, the groundwork is being laid for a more comprehensive evaluation system that may survive after Bank financing ceases.

John Raven takes the case studies as grist for a broadbush analysis of the role of evaluation on what he calls a managed economy. He discusses the need for information in an economy where most investment decisions are
substantially divorced from market forces and argues that the evaluator is most useful in the role of policy researcher. His paper explores the implications of this view for evaluation methodology, the institutional framework for evaluation work, the role of consultants, and the professional development of staff. In particular, he urges the need for evaluators, in contrast to academic researchers, to go beyond their data, to speculate, and to apply their experience and knowledge to the task of interpreting partial and preliminary findings.

The four authors represented in the volume were chosen not only for their professional credentials but deliberately because of their direct involvement with the projects under review. Middleton served as a consultant to the Philippine Radio Education Project during its early years, working with the curriculum developers on formative evaluation. Schwarz, as noted above, was a Bank consultant who visited Pakistan several times to work with the Pakistanis on the evaluation component. Sack, also a Bank consultant, has worked on project preparation and appraisal, as well as supervision, for several of the projects supporting the Haiti educational reform. Finally, Raven was one of the consultants under contract to the Government of Pakistan to assist in implementing the evaluation component of the Primary Education Project. Choosing as authors people who were closely involved with the projects was deliberate and represents a judgement that the advantages of prior knowledge of people and circumstances outweighs the possible pitfall of loss of objectivity. It is the editor's view that the papers presented here, rich in detail,
expressing in some cases contrasting points of view, and replete with lessons, amply vindicate the strategy of using insiders as illuminators.

Barbara Searle
The World Bank
Education and Training Department
SECTION I: 
THE PHILIPPINE RADIO EDUCATION PILOT PROJECT:
A CASE STUDY OF DECISION, DESIGN AND INTERACTIONS

Dr. John Middleton, Director
School for International Training
Brattleboro, Vermont USA
Preface

Opportunities to delve deeply into complex projects of educational innovation are rare. I have been fortunate to have been asked by the World Bank to conduct such a study of a radio education pilot project in the Philippines funded by the Bank and the Philippines government. To have this opportunity for a project with which I worked as a consultant and which I admire greatly is a significant challenge.

It is not easy to untangle the many threads of thought and action out of which the Philippine radio education pilot project was woven. And yet, the project was such a significant endeavor that the effort is important. Hopefully, by helping the key actors in the project tell their story, by integrating the various perspectives from which the project was viewed, and by adding my own analysis I have provided a case which will help others in designing and carrying out future pilot studies.

The study has been developed from a variety of data sources. Interviews were conducted in the Philippines in January, 1984. Key bank staff were interviewed at various times during the period December, 1983 to June, 1984. Project files, both in the Bank and in the Philippines, were examined, as were principal project documents. The two months I spent as a consultant to the project in the Philippines in 1978 and 1979 provided an invaluable personal and professional experience, as well as an immediate sense of the project and its leaders which (for better or worse) informed the current study.

Early drafts of the study were reviewed by individuals in the Bank and in the Philippines who worked closely with the project. Their commentary has been invaluable, and has contributed significantly to the study.

I am grateful to Dr. Barbara Searle of the Bank who saw the need for this study and provided significant personal and professional encouragement as it was being developed. Other Bank staff, notably Shigenari Futagami, contributed significantly. My colleagues in the Philippines were open and frank in helping me tell their story. My respect for them and their work is great; my gratitude for their help with the study greater yet. Of course, whatever part these individuals have had in developing the study, the ultimate responsibility is mine alone.

John Middleton
Brattleboro, Vermont

December 31, 1984
A NOTE ON NAMES

Whether or not to name the characters in a case study is always a difficult question. In most instances, real names are not important to learning from the case. Not using real names is usually the most comfortable solution for the key actors in a case, although this is not always true. The absence of real names, however, can seem to reduce to versimilitude of the study.

Given the evaluative nature of this case, it has been decided not to use personal names in the narrative. This, in turn, has made it necessary to create descriptive names for the key actors in the study. Because made-up names are more difficult for the reader to internalize, these "names" are listed below together with a brief description of the individual role.

BANK STAFF

**Evaluation Expert**— One of two Bank professional staff deeply involved in the design and supervision of the project. An economist and leading scholar in educational technology for developing countries and in evaluation.

**Mass Media Specialist**— The second of two Bank professional staff continuously and directly involved in project design and supervision. A specialist in broadcast education.

**Appraisal Team Leader**— An economist in the Bank's East Asia Projects Division who led pre-appraisal and appraisal missions for the large educational development project towards which the radio pilot was addressed.

PHILIPPINE STAFF

**Pilot Project Director**— Director of the Educational Communications Office (ECO) which had overall responsibility for the pilot; had participated in early study phases as well. A mass communication specialist.

**Evaluation Consultant**— A part-time consultant on summative evaluation to the Pilot Project Director. Primary position was in the Ministry of Education.

**Evaluation Director**— A senior mass communication research scholar and leading participant in the study phases of the project. Directed the pilot summative evaluation under contract to ECO.

**CTRE Director**— Another senior communication professor. Directed the rural education component of the pilot project under contract to ECO.
Introduction

"If the pilot project proves effective, its successive expansion would be considered as a part of a possible project for educational materials development. The pilot project includes a substantial evaluation component, timed to provide information on cost and effectiveness in time to be of use in deliberations concerning the magnitude and composition of a possible educational materials project." World Bank Memo, 9/30/77

"Perhaps the most important output of the project is information on the comparative cost-effectiveness of several of the relatively few potentially alternative approaches to cost-effective improvement of quality in primary education." World Bank Review Mission report, 11/77.

"Materials developed under the project are being used on a national scale ... In this respect, the project has had a major impact on national education. Mass media, however, remains a supplemental activity and does not play a major role in education."

"Experience gained under this project suggests that the Bank is not the ideal agency to carry out small-scale, innovative projects."

--- World Bank Project Completion Report, 8/83

In 1978 the Government of the Philippines embarked on a three-year pilot project to test the use of communication technology in education. The project was well supported by the Philippine Government and a loan from the World Bank, and Bank staff played significant roles in the design and supervision of the project. A substantial field experimental evaluation was built into the project. As the planning documents make clear, this evaluation was intended to produce information for decision making in 1980-81 on further use or expansion of communication technology as part of a subsequent World Bank education loan.
The pilot project and the evaluation were completed on schedule. The project generated a large body of materials and reached all utilization targets. The evaluation was completed as designed, and reports were produced on time. The evaluation data indicated that the pilot project led to modest learning gains, and provided insight into the relative cost-effectiveness of various combinations of the program elements included in the study. The effort was completed well within budget, with approximately 40% of the available funds unspent.

And yet, the project was not continued, much less expanded. Moreover, it seemed that the extensive evaluation data had little to do with final decisions. It was not clear how these decisions were reached. Despite successful implementation, the project had not succeeded in its principal purpose of supporting rational decision-making. The validity of Bank involvement with such efforts was questioned.

What happened? The pilot project had more than adequate funding, was effectively managed, and was subject to a sophisticated field evaluation. Clearly, the answer to the question is of considerable interest, not only to the Bank in terms of its role in such endeavors, but also to the many development professionals and organizations considering the use of pilot projects to inform decisions.

The present case study was commissioned by the Bank to "...draw lessons that will help improve the design and implementation of evaluation components in education projects in the future." The study is to "...attempt to explain the way in which evaluation functioned in the Educational Radio Project: how it was conceptualized and designed, how it was implemented, and how the results were used in decision-making."

The answers that emerge in the study are not simple. Understanding the complex process through which project outcomes developed requires integration of the perspectives of the key actors, perspectives which changed as the project was carried out.

Thus a story is needed, one which seeks to show how the process developed and unfolded over time. The story, in turn, must be subjected to an analysis which seeks to draw out and support with data the principal patterns which help us understand the process, and which lead to the insight that strengthens future use of pilot projects in development.

The first four chapters of this study present a narrative description of the pilot project. The emphasis is on those aspects of the project most clearly connected with the evaluation and decision-making processes. Thus it is not a complete project history. Many interesting and valuable aspects of the pilot receive only cursory treatment — the materials design and development process, the challenges of implementation, and the achievement of educational goals and objectives. These aspects are well documented elsewhere. 1/

The patterns and lessons of the case may be found in Chapter Five, together with a summary of the implications of the case for future Bank involvement in educational pilot projects.

A note on the methodology of the study may be found in the appendices.
Chapter One: Project Initiation

As the 1970’s began the Philippine government initiated a decade of concentrated effort to improve education. A number of important milestones in educational development had already been achieved. Enrollment rates were high, with virtually all children entering primary school, and 60% completing grade six. Twenty-one percent of the relevant age group was enrolled in higher education. Males and females benefited equally from educational opportunity, and adult literacy stood at 90%.

Serious challenges remained. Repetition rates were high, reducing the efficiency of schooling. There was serious inequity of access between urban and rural populations.

The overriding problem was one of quality of education. There was, on the average, one textbook for each ten students. While student/teacher ratios were favorable, teachers had limited opportunity for in-service education. The curricula of the primary and secondary schools needed reform, and there were serious concerns about the quality of much higher education.

The problem of educational quality was most serious in rural areas. Rural primary and secondary schools were weak. And despite a widespread non-formal education effort, a large rural population of out-of-school youth and adults lacked opportunities for basic, agricultural and vocational education.

With Bank support, the government addressed a number of these problems. Higher education and teacher-training institutions were strengthened. A major effort was launched to revise curricula and significantly increase the availability of textbooks at the primary level.

The large-scale program of educational development led to the creation of a specialized agency of the Ministry of Education and Culture (MEC) to develop and manage foreign assisted projects, including those funded by World Bank loans. Created in 1970, the Educational Development Projects Implementation Task Force (EDPITAF) was chartered for a ten year period. EDPITAF rapidly assumed a vital leadership role in the development of Philippine education.

COMMUNICATION TECHNOLOGY: THE STUDY PHASE

In 1975, EDPITAF turned its attention to the potential of communication technology to improve the quality of education. The Philippines had used both radio and television for education, but lack of broadcasting infrastructure and other problems had led to the abandonment of the effort. Nevertheless, evaluation results showed that broadcast education could improve quality. In addition, a number of other countries were successfully using communication technology, notably radio, for educational
purposes. In particular, a primary radio mathematics project in Nicaragua was evolving effective curriculum and materials design procedures and demonstrating significant impact on pupil performance at very low cost. World-wide interest in the use of satellite communication for development was also increasing, and some leaders in the Philippine government were interested in this aspect of the technology.

EDPITAF received Bank support in early 1975 for a Preinvestment Study of Communication Technology for Education (PIS-CTE). Part of a package of initiatives for continued collaboration with the Bank in educational development, the study was broadly conceived to include both formal and nonformal education. It would encompass the use of satellites, and would focus on priority goals for improving the quality of rural education. The study was clearly seen as leading towards proposals to the Bank for incorporating communication technology into the educational development effort.

The proposed study was supported at the highest levels of government, including the President's Office and the National Economic Development Authority (NEDA). An inter-agency planning committee (IAP) was formed by executive order from the Office of the President of the Philippines. It was chaired by the Assistant Executive Director of the President's Executive Office, and included all relevant agencies: The Department of Local Government and Community Development, MEC-EDPITAF, the Department of Public Information and its National Media Production Center, the Department of Public Works and Telecommunications, the National Economic Development Authority, the Philippine Communication Satellite Corporation, the Broadcast Media Council, the Rural Broadcasters Association, and a senior communication professor from the University of the Philippines (UP), representing schools and universities.

PIS-CIT Workshop

The interagency planning committee convened a workshop in June, 1975 to review and establish the terms of reference for the study. The workshop received funding support from EDPITAF, the United Nations Development Programme (UNDP), and UNESCO.

The workshop brought together senior education and communication leaders from the Philippines. International experience was brought to bear by a team of five consultants sponsored by UNESCO.

Appearing in the participant list are a number of individuals who later went on to play key roles in the pilot project. These included the Director of the UP Institute of Mass Communication, a member of the IAP and later director of the pilot summative evaluation; the eventual Pilot Project Director, who served in the workshop secretariat; and a professor of development communication at the University of the Philippines—Los Banos, who would direct the Rural Education component of the pilot. The then EDPITAF Executive Director, a strong supporter of communication technology, played a leading role.
The international team also included individuals who would have continuing association with the project. A leading economist and evaluation specialist, then of the Educational Testing Service, would move to the World Bank and play a central role in project design, evaluation and supervision. Other consultants came from the US Department of Health Education and Welfare, the British Open University and the Center for the Production for Television for Adults in Singapore. These would later serve as consultants to the pilot. A senior mass media consultant from UNESCO rounded out the team.

The workshop was enriched by the preparation of several substantive papers. These included inventories of educational and communication policies and goals for the country, as well as an inventory of broadcasting resources.

By all accounts, the workshop was an extremely positive and successful endeavor. The PIS-CTE study was framed and terms of reference developed, including a rather complete management plan and budget. The study was to begin in January 1976 and be completed at the end of the same year.

Some excerpts from the terms of reference capture the direction of the study:

"The government envisages the use of satellite communication technology for education purposes, both in the formal and non-formal sectors.... To provide the government with a sound basis for its decision-making process, and to assist in selecting the most effective strategy for introducing and developing educational technology on a nation wide scale, it is proposed, as a first step, to provide a pre-investment study....it is expected that the Government will be able to decide on the feasibility of the massive introduction of advanced educational technology, using satellite communication, and to choose the priority point of entry into the educational system."

The thrust is clear: a major study as the basis for national decisions on a national system of educational use of communication technology, with special attention to the issues regarding satellite communication.

The study plan specified a series of interim reports, and that the final report would be submitted to the President's Office for review and decision. Provision was made to assemble an expert team, and to maintain the involvement of key agencies throughout the study period. EDUITF was selected as the coordinating agency responsible for the implementation.

The Study

The first phase of the study was devoted to assessing needs for educational development and setting objectives for the uses of technology. Four primary needs were identified:

1. promoting more effective learning
2. increasing the relevancy of what was learned
3. reducing the cost of education
4. serving a large population in school and out of school

This stage also resulted in the development of a base of information on Philippine education and broadcasting.

Importantly, radio was identified at this early point as the medium of choice for the Philippines, given the existing availability of broadcast resources and the nature of identified needs. The study, and the eventual project, would build from existing media resources.

Eight project options were developed in the second stage of analysis. These were strategies with potential to meet identified needs. Six of the eight options would support non-formal education. Two dealt with formal education, one for teacher training and the other for quality improvement in elementary education. 5/

Recognizing the large scope of these eight options, the study team consulted at this point with the Interagency Planning Committee.

Narrowing the Options

The IAP stressed the difficulty of carrying out all eight options. Guidance from the top was clearly needed, and advice was sought from the Office of the President in January of 1977. The response was clear: proceed to develop options in two of the eight areas, basic (primary) education and rural development.

The final stage of the study saw the production of a five-volume project proposal to establish the National Educational Communication Program. The Program was designed to meet the directive to support primary education and rural development, and rested on six strategy elements: b/

1. Development of a teaching approach that integrated interpersonal and multi media communication, primarily radio. This approach would be followed in both formal and non-formal settings.

2. Decentralization of educational communication activity. Regional production and coordination centers, as well as a monitoring and feedback system would be developed.

3. Establishment of a central office to manage the project. This would be primarily a coordinating unit.

4. Developing collaborative relationships with education and development institutions. This would be a primary mechanism for mobilizing existing resources and expertise while maximizing institutional development.
5. Utilization of existing production, dissemination and research resources. The study had demonstrated that adequate resources already existed.

6. Adoption of a technical support design. The design relied heavily on use of existing facilities, with new acquisition as needed to supplement current system resources.

These elements added up to a national educational communications system addressed to needs in primary and rural education, with significant decentralization.

**Decision: The Pilot Project Is Born**

The proposal was submitted to the National Economic Development Authority, the National Board of Education and the World Bank.

The decision from NEDA came in August, 1977: 

"1. The pre-investment report does not demonstrate the market, technical and economic feasibility of the project ... It is, therefore, recommended that the COMITECH be considered on an experimental/pilot basis to test its feasibility from all standpoints...

2. It is further recommended that the project be redesigned into a smaller version consisting of two or three communication centers coupled with a few earth stations to service a more limited area."

After more than two years of intensive study for a national system, the proponents of educational communication had gained the opportunity for a pilot project. What had begun as a satellite-based national system had, through the process of policy research and decision, become a limited scale exploratory effort. The focus still was on primary and rural education.

Yet despite the sharp reduction in scale, the effort had led to action. Moreover, there was a base of data, analysis and committed persons coming out of the study. All would be useful as the pilot project moved ahead.

**ENTER THE BANK**

With the government's feasibility study completed, and fundamental policy decisions reached, the Bank sent a team to the Philippines in September, 1977 to appraise plans for the pilot project developed by EDPITAF. These plans called for the use of radio to upgrade in-service primary teachers, for direct primary classroom instruction and to improve rural education. Radio would be used in conjunction with interpersonal communication and printed materials.
The team consisted of a Mass Media Specialist in the Bank's Education Department, and the Evaluation Expert, who had participated in the PIS-CTE Workshop, later joining the Bank's Research Department. Both would have significant continuing contact with the project throughout its life.

The Mass Media Specialist had been with the Bank for fifteen years, concentrating on the use of mass media for education. He was one of a very few persons in the Bank with this expertise, as communication technology had not been a Bank priority.

The Evaluation Expert was an economist who had specialized in educational technology in developing countries. He had written several of the standard books and articles in the field, and brought wide knowledge of the use of technology in education to the project, along with significant experience with research and educational evaluation. As a participant in the PIS-CTE workshop, he knew the key persons in the Philippines well. His experience and expertise were well-respected, and he was seen by Philippine counterparts as an ideal partner in the enterprise.

It was, in some respect, an unusual Bank team. The project would be funded as part of the program of the East Asia Education Projects Department, and would normally be appraised and managed from that office. The technical requirements of the project, however, as well as its small dollar value, led to the fielding of this particular team.

PILOT PROJECT DESIGN

The Bank team worked with Philippine counterparts for two weeks to develop a proposal for the pilot project. The design developed during this period grew out of the proposals prepared by EDPITAF in August, and benefitted from the extensive analysis and planning that had been done under the pre-investment study. It became the basis for Bank funding and was followed, with only minor variations, for the full period of the pilot project.

The document that was developed was not the usual Bank project appraisal report, although it would serve as the basis for the project loan. It was a plan for the project jointly developed by Philippine experts and Bank staff — not a Bank analysis of a plan prepared by Philippine leadership.

The pilot project that emerged focused on the use of radio to improve the quality of education. Two primary uses of radio were to be tested: 1) for in-service teacher training and 2) for direct instruction in language arts in rural primary schools. There were two additional components. One was an exploratory study of the use of radio to support rural non-formal education; the other was the study of a number of technical issues associated with the use of technology in education. Thus the project had four action components.
The design also specified a comprehensive set of evaluation activities. There would be extensive formative evaluation of materials as the project developed. Large-scale summative evaluation would be conducted to assess the costs and effects of the two instructional components -- teacher in-service education and direct primary instruction.

At the time that the design was being developed, the Government and the Bank had initiated a national project to greatly strengthen the quality and availability of textbooks in the schools. There was considerable interest in the potential interactive effects of the new textbooks and the radio project in improving educational quality. At the suggestion of the Bank Evaluation Expert the summative evaluation design incorporated measurement of the effects of various combinations of the two radio components with different ratios of new textbook availability to students. 9/

As is the case with many pilot projects, there were several objectives. World Bank documents highlighted two. One, of course, was to develop research-based information on the costs and effectiveness of radio in improving the quality of primary education. In addition, Bank objectives included exploration of the use of radio for rural education, as well as "... continued examination of the technical options for production, interconnection, transmission and reception that will be open to the Philippines should the government subsequently decide on a major expansion of its use of communication technology for education." 10/ These objectives centered on learning from the project to support future decisions.

EDPITAF shared these objectives. There were two additional anticipated outputs, however, on the Philippines side. One was service: the project was expected to have beneficial effects for teachers, students and rural people. A second was institution building: the project would create a "nucleus of management, coordination, [and] operations ... resources" useful in establishing an overall national system. 11/ Project managers at EDPITAF also placed considerable emphasis on developing "instructional models" in each of the applications areas.

The action project was complex. A sense of the many components, and their interrelationships, is important to understanding the processes of implementation, evaluation and decision-making. A summary is presented below; more detailed analyses are available from various project documents.

**Continuing Education for Teachers (CET)**

The in-service teacher education component was known as Continuing Education for Teachers (CET). Semester-length courses would be provided by radio for primary school teachers in Pilipino (the national language), English, social studies, science and mathematics. Subsequent to the original design, an additional course on teaching skills, called Educational Trends, was added.
The courses would integrate broadcast radio, print materials and periodic forum meetings. Teachers could receive in-service or academic credit for successful completion of other courses. Regional colleges and universities would cooperate in the forums and in granting credit; radio lessons and print materials would be developed centrally. Broadcasting would be through regional stations.

Courses would be developed and used on a staggered schedule. Pilipino would come first, with subsequent courses being added as the project progressed. The Educational Trends course was subsequently added to Pilipino for the first (formative) year of operations.

It was anticipated that more than 12,000 teachers would participate in the program.

Radio-Assisted Teaching in Elementary Schools (RATES)

The direct classroom instruction component was known as Radio Assisted Teaching in Elementary Schools (RATES). RATES was designed to improve student abilities in Pilipino. In a nation of many languages, achieving widespread effective use of Pilipino had high priority.

RATES was to be the built around daily 30-minute radio lessons, supported by teacher's guides and pupil worksheets. As with CET, materials would be developed centrally by EDPITAF, and broadcast/disseminated at the regional level. Initial plans called for the development of 180 lessons per year; subsequent adjustments to the reality of the classroom schedule reduced this to 140 per year.

The first year of RATES would be directed at Grade IV pupils. Grade V materials would be developed for use in the second year; and grade VI added in year three. It was expected that approximately 2,500 classes would be reached with RATES materials.

RATES lessons were to be developed following the official curriculum of the Department of Education, insuring consistency with the scope and sequence of teaching and with the new textbooks.

Communication Technology for Rural Education (CTRE)

Listed as a special study in the design plan, CTRE would evolve into a full test of the use of radio to support rural extension education. CTRE would be coordinated with another Bank supported project, the Philippine Training Center for Rural Development (PTCRD). It would use radio to support person-to-person contacts and extension groups.

Technological Studies

The pilot project would provide the opportunity for a variety of technical studies on production and broadcasting. A principal option to be explored would be the establishment of a satellite-based communication link between EDPITAF and a regional education office to facilitate decentralized educational administration.
SUMMATIVE EVALUATION DESIGN

Given the principal purpose of the project, the summative evaluation was of great importance, and was designed in some detail.

The evaluation was designed to answer this basic question: "How do CET, RATES, and additional textbook availability — introduced separately or in combination — affect student performance?" 12/

To answer this question, a multi-treatment field experiment was designed. There would be eight "treatment" combinations of RATES, CET and different textbook ratios. The ninth treatment would be a control. Classrooms were randomly assigned to treatment groups. The primary dependent variable would be student learning gains as measured with pre- and post-tests.

The design included provision for a cost analysis of the various treatment combinations.

The period from the beginning of the project in April, 1978 to June 1979 was set aside for detailed development of the evaluation effort, including sampling plans, development and testing of instruments, and try-out of procedures. The timing would permit a preliminary evaluation based on data collected during the tryout phase, both of the action programs and the evaluation instruments and procedures. This period coincided with the formative year of implementation of RATES and CET.

The "pilot" evaluation would be conducted during the second year of the project, June, 1979 to September, 1980. The report would be available in September in time for consideration in planning for the anticipated Bank-funded education project.

Third year evaluation activities would focus on fifth grade students with two years of exposure to the various treatments, and on other activities to be specified.

PROJECT SITES

The purpose of the pilot was to test impact on the quality of rural education. Accordingly, pilot sites were selected which met the definition of "rural." These were the provinces of Pangasinan and Leyte. In both provinces significant percentages of the population did not speak Pilipino (45% and 72% respectively), providing a good test of the impact of the Pilipino courses.

Pangasinan is located approximately 100 miles north of Manila, where EDPITAF project headquarters were located. Leyte is an island province located approximately 500 miles to the southeast of Manila, requiring air travel. The regional education offices in both provinces promised full support and cooperation in carrying out the project and the evaluation.
PROJECT ORGANIZATION

As designed the pilot project was large and complex. Moreover, there was little if any room for flexibility in time schedules. The Bank Review Mission Report — effectively the plan for the pilot — mandated that the project be implemented and evaluated on time in order to meet deadlines for decisions regarding the upcoming national education project, anticipated to include a national educational materials development corporation.* A strong and effective project management structure was clearly needed.

Division of Labor

It was also clear that the resources of EDPITAF needed to be augmented if the pilot were to be successful. There was also a desire to maintain the overall government strategy of involving key organizations. Consequently, the work of carrying out the project was shared among several agencies.

The Educational Communications Office (ECO) was established within EDPITAF. ECO would have overall responsibility for project management, and specific responsibility for implementing RATES, CET and the studies of technological options. ECO would be led by the Pilot Project Director.

The rural education component of the project (CTRE) would be carried out by the Department of Development Communication of the University of the Philippines, Los Banos (UPLB). A senior professor at the university, who had also participated in the pre-investment study, would be CTRE Director. As the national agricultural university, UPLB was active in extension and rural education, and was the home of the PTCRD project. ECO would provide overall supervision of this component, which would be implemented under a sub-contract.

The summative evaluation, including the evaluation of the CTRE component, would be conducted by the University of the Philippines, Diliman. It would be led by the Dean of the Institute of Mass Communication, also a member of the Interagency Planning Committee. The Department of Economics would conduct the cost analyses. ECO would provide overall administration; the work would be performed under a contract between ECO and IMC.

IMC was the logical choice to conduct the evaluation. It was perhaps the leading communication research organization in Asia, and had a long record of success in field research and evaluation.

* This project eventually became a $100 million educational sector loan known as the Program for the Decentralization of Educational development (PRODED).
**Linkages**

ECO would establish linkages with other agencies. Of particular importance was the Bureau of Elementary Education of the Ministry of Education and Culture. Responsible for primary education, BEE would play a key role in implementation of the project, and would be the natural home for institutionalization if the pilot succeeded. Moreover, the technical expertise of BEE would be essential in the development of CET and RATES materials.

Linkages were formalized through the establishment of a Project Advisory Group consisting of four consultants to the project. One of these was the Pilot Project Director. The others were the IMC Evaluation Director; the Head of the Bureau of Elementary Education; and the Vice-president of the Philippine domestic satellite corporation.

The PAG was to serve several purposes. One was to provide overall policy guidance for the project. During the first two years of the pilot it would also advise on operational issues. Perhaps most importantly, the PAG could serve as a linking mechanism between the pilot project and higher level decision makers. The Head of the Bureau of Elementary Education, for example, was an important potential link to the Ministry of Education and Culture. The BEE was providing technical consultation to ECO on curriculum development, and was a potential home for radio education once the pilot phase was finished.

The broad organizational structure of the pilot project is shown in Figure 2.

**Negotiation and Funding**

The Mass Media Specialist and the Evaluation Expert returned to Washington, D.C. at the end of September, 1977; the full report of their mission, which incorporated the design of the project, was completed in mid-November. The Bank's East Asia Education Projects Division, which would have formal project supervision responsibility, approved. The Philippine National Board of Education concurred in December of 1977. The loan was negotiated in Washington in February of 1978, and received Bank approval as part of the Fifth Education Loan in April. The National Economic Development Agency gave formal approval in July, and the project was officially begun in August, 1978.
Figure 2: Project Organizational Structure
The total value of the project was estimated at US$3.86 million, of which $2.0 million would be foreign exchange costs funded under the Bank loan. Bank funding would provide a number of critical inputs:

**Equipment**

1. Vehicles
2. A recording studio for ECO
3. Radio/cassette players for schools

**Technical Assistance**

1. Twenty-two person months of consultancy for RATES and CET;
2. Twenty-four person months of consultancy for the summative evaluation;
3. Sixteen additional person months for rural education and other studies;

**Fellowships:** Twenty-four months of fellowships for study abroad for the project as a whole.

Government counterpart funds would establish ECO and fund local costs of implementation and evaluation.
Chapter Two: Formative Year

The pilot project officially began in August, 1978, with the completion of all necessary legal steps for approval of the Bank loan within the Philippine government. Work actually started much earlier in anticipation of approval, and with funding support from EDPITAF and the balance of Bank funds remaining from the pre-investment study. The organizations collaborating in the project were in place and had been working on implementation for several months. The project was begun with considerable energy and commitment.

IMPLEMENTATION SUMMARY

Broadly stated, the objective for the formative year was to develop and test the full range of mechanisms necessary for the pilot test of the project in the next year.

For ECO, this meant designing procedures for materials development; developing prototype materials and conducting formative evaluation; establishing field offices and operations in the pilot sites; and implementing both CET and RATES in the process.

ECO also confronted a number of difficult administrative challenges. With contracts in place, effective working relationships with the two principal contractors — UPLB and IMC — had to be developed. Linkages needed to be established with Education and Culture and with other cooperating agencies. The resource base of the project — tape players, vehicles, a recording studio — had to be put in place. The Bank had provided funding for technical consultants and a fellowship program for external training of key staff: consultants needed to be identified and fielded, and staff selected and sent for training. It promised to be, and was, a busy year.

It would be an equally busy year for IMC staff. The complex summative evaluation plan had to be actualized. Instruments and procedures had to be developed and tested as RATES and CET were fielded in prototype form, leading both to a preliminary evaluation report to guide ECO in the pilot year and to the establishment of an effective, full-scale pilot year evaluation effort.

At Los Banos, the CTRE staff needed to develop project linkages with the Philippine Training Center for Rural Development project. It was through PTC that CTRE would establish links with project clientele. The same materials development and operations procedures challenges that confronted ECO faced CTRE as well.
The main tasks — getting CET and RATES in operation, testing and developing summative evaluation procedures — were accomplished well and on time. CTRE found the proposed linkages with PTC to be impractical for a number of reasons, modified the project design as necessary and proceeded with the task of developing a test for radio for rural non-formal education. As noted in a Bank supervision mission report in January, 1979 the project was being carried out "... without major difficulty under the efficient management of ECO." 1/

And yet there were problems and patterns which, in retrospect, foreshadowed later developments that significantly affected the use of the evaluation in decision-making. These related primarily to RATES, CET and the summative evaluation itself.

Implementation and Evaluation: Establishing Relationships

The project design did not provide much detailed guidance on the ways in which implementation would be linked with summative evaluation. While it was clear that IMC would conduct an independent evaluation, it was also clear that ECO had overall responsibility for the full pilot project. This relationship was reflected in the contractual arrangements between ECO and IMC, which made provision for a consultative and reporting relationship. Effective management of the sequence of implementation and evaluation activities required no less. At the same time, it was recognized that the expertise for evaluation was in IMC. The way in which the two organizations worked together would be important.

ECO appointed a summative evaluation consultant, of the Planning Office of the MEC, to advise the ECO director on summative evaluation. This not only provided an important level of technical competence within ECO, but also established a potentially valuable link between EDPITAF/ECO and the Ministry.

The Evaluation Consultant's role included review, on behalf of ECO, of various elements of the summative evaluation design and instruments being developed by IMC. In October, 1978, a meeting was held between ECO and IMC to go over the evaluation design and instruments. ECO brought a number of concerns to the meeting. 2/ One had to do with the cost analysis. While the design made provision for comparative analysis of the costs of the various combinations of RATES, CET and textbooks under study, there was apparently no plan to gather cost data on current teaching of Pilipino as reflected in the "no treatment" control groups. Thus while it would be possible to make cost-effectiveness comparisons among treatments, it would not be possible to compare the costs of any treatment with the costs of current teaching. This, it was thought, would make decisions on adoption of any of the pilot models difficult.

There were also concerns regarding the learning-gain tests to be used to assess the impact of the various treatments. The tests had eighty items, ten of which measured listening skills. The balance assessed reading competencies: ECO felt that these tests would not be sensitive to the primary impact of radio instruction, and urged a better balance between listening and reading skill items.
The Evaluation Director was not present at the meeting. However, it is reported that IMC staff did not respond favorably to these criticisms. It is a matter of record that the design was not modified on either count.

From ECO's point of view, this meeting marked an important turning point in the pattern of working relationships between ECO and IMC. After the meeting, relationships became less collaborative. The IMC continued to submit progress reports. But evaluation instruments and documentation were shared with ECO only after they had been adopted for use, effectively cutting ECO out of the process of evolving the evaluation system.

**Logistics**

At the time the meeting occurred, IMC was trying unsuccessfully to obtain permission from ECO to purchase a vehicle — a critical resource in enabling IMC staff to carry out their responsibilities in field sites. EDPITAF was unable to give permission because of a government-wide restriction on vehicle procurement. This ban would last the entire life of the project, preventing EDPITAF from buying the seven vehicles budgeted in the Bank loan.

The impact was felt at both agencies. The lack of project vehicles forced reliance on already over-burdened car pools. Great ingenuity was shown in finding alternative ways to deliver materials, conduct formative evaluation and coordinate project activities in the field. However, a great price was also paid in terms of diversion of staff time from substantive to logistical activity. Contact with field-level activity also suffered.

Procurement of vehicles was only one of several problems encountered in establishing the physical resource base of the project. Philippine government and Bank international competitive procurement procedures also delayed the acquisition of cassette players and recording/duplicating equipment. Tape players eventually arrived four months later than planned. Requirements for acceptance of lowest cost bid led to the purchase of equipment of lower quality than desired, which in turn created maintenance problems. Delays in acquiring tape duplication equipment adversely affected the quality of radio materials and, until the equipment arrived in 1980, made duplication a slow and cumbersome process.

The Bank had also funded the construction of a recording studio in EDPITAF facilities in Manila. Audio materials were produced in rented facilities while the studio was being constructed. After it was completed in late 1978, it was found to be inadequately insulated for sound and poorly air conditioned. As a result, primary production continued in external facilities, although the studio was eventually useful for editing and duplication.
Technical Assistance and Training

The Bank had funded technical consultancy services for both ECO and IMC. In addition, EDPTAF had access to funds left over from the pre-investment study. Consultant services were expected to provide an important level of international expert support for the project. This would facilitate incorporation of lessons learned regarding implementation and evaluation from radio education projects elsewhere.

ECO began using external consultants almost immediately to assist in curriculum and materials development procedures, in design of formative evaluation systems and in management planning. IMC, on the other hand, felt that external technical assistance was not necessary. IMC staff were very well qualified to conduct empirical communication research. Moreover, the Bank had provided important technical consultation through its Evaluation Expert. During the formative year IMC used only two weeks of its external consultancy resources — for advice on test construction.

The Bank Evaluation Expert visited the Philippines on mission in January, 1979, and worked closely with IMC staff in further development of the evaluation design. As noted in the report of this mission 4/ (the team was comprised of the Evaluation Expert and the Mass Media Specialist), the summative evaluation was "... proceeding smoothly." This mission reaffirmed the purpose of the pilot project: "If the study gives any encouraging indications as to the effectiveness of primary level instructional radio ... it would be worthy of Bank financing to follow up these projects for developing primary education region by region with new textbooks and educational radio along with improved curriculum."

The technical collaboration between the Evaluation Expert and IMC was a strong component of the development of the evaluation during the formative year. In addition to the mission in January 1979, the Evaluation Expert corresponded directly with the Evaluation Director on technical problems associated with the evaluation. Together, they planned a book on the project which would include, in addition to the evaluation data, technical chapters on various aspects of the pilot effort.

During the formative year, ECO sought to make use of the fellowship funds made available by the Bank. Once again, Philippine government regulations posed barriers. Government rules restricted fellowship opportunities for study abroad to permanent staff. As ECO staff were all on a contract basis for the term of the pilot, they were not eligible to take advantage of the opportunities provided. Although some staff would eventually benefit from study tours, the fellowship component was never fully utilized. Efforts to obtain exceptions to the rules, moreover, placed yet another administrative burden on ECO leadership.
ACCOMPLISHMENTS IN THE FORMATIVE YEAR

Despite the difficulties encountered, the project as a whole went forward largely on schedule. ECO successfully established materials development and formative evaluation procedures. Audio and print materials were developed for both RATES and CET and distributed to pilot sites for utilization. In the formative year, cassette tapes were used for audio materials in the classrooms. IMC completed planning and instrument development for the summative evaluation. Data collected during the formative year was analyzed and shared with ECO and the Bank.

Preliminary Evaluation Findings

The preliminary evaluation report prepared by IMC (based on data collected during field trials of procedures and instruments) provided useful insights for the pilot year. In combination with formative evaluation data collected by ECO, it pointed towards improvements in the learning materials and utilization systems. Problems encountered in test administration were identified. Importantly, the preliminary evaluation report noted that teachers in control areas were exerting special effort, apparently to "win" in competition with RATES classes. This was held to partially explain the finding of "no significant difference" between RATES and control classes during the formative year.

IMC submitted its preliminary evaluation report in December 1979, well into the implementation of the full pilot test. The extent to which the findings were shared with ECO during the year is not clear. The general pattern of interaction between ECO and IMC at the time would indicate that information did not flow as freely as might have been hoped.

Modification of Summative Evaluation Plans

As a result of the Bank mission in January, a number of modifications were made in the summative evaluation plan. At the request of EDPITAF, an additional pilot site was added in Cotabato (Mindanao). This site would add a Muslim area to the pilot test. A separate evaluation would be conducted by Notre Dame University, focused only on RATES and CET (without textbooks). It was agreed that the major effort in the pilot year would be in the Pangasinan site. Both RATES and CET would be expanded there, while the project in Leyte would be smaller. The dates for the pilot evaluation were confirmed as coinciding with the 1979-80 school year (June to March).

Planning for the Future

As the formative year unfolded the various organizations involved in the effort maintained focus on the use of the pilot for future decisions. The Bank mission in early 1979 had centered attention on the summative evaluation. In keeping with the emerging Philippine strategy for educational development in general, EDPITAF was beginning to think about using the pilot project model for educationally disadvantaged regions.
Planning for utilization of the summative evaluation was not a high priority at this time. The challenges of getting the pilot year underway loomed large and occupied center stage. Awareness of the impending decisions, however, was high.

THE ROLE OF THE BANK

Bank staff continued to play an important role in the project during the formative year. Two missions were completed. In the first, the Mass Media Specialist came alone in May 1978 as the project was getting underway. He reported that preparations were going smoothly, and that the EDPITAF Director had raised "... the question on how to follow-up the present pilot by a gradual nation-wide application of educational radio if it is proved effective." 6/

The second mission took place in late January and early February of 1979, involving both the Mass Media Specialist and the Evaluation Expert. As noted earlier, this mission was important to modification in the summative evaluation plan. The report recognizes these changes, notes that the summative evaluation was proceeding smoothly, and discusses EDPITAF interests in expansion of the pilot as part of the forthcoming project to strengthen basic education. The need for thorough examination of the cost and logistical implications of expansion was noted.

The Mass Media Specialist and the Evaluation Expert had become the Bank's supervision team for the project, although the Bank's East Asia Education Projects Division had formal supervision responsibility. The official project officer was a staff member of that Division. With the other staff conducting the field missions, the East Asia Division staff member had no direct field role in the project. The project's supervisors and potential advocates were not in the division where decisions would be made, not only about the pilot, but also about the future education project toward which the pilot was aimed.
Chapter Three: Pilot Year and Beyond

The formative year had been completed. In the period June, 1979 to June, 1980, ECO would implement RATES and CET in three sites. IMC would carry out the summative evaluation in two of these sites (Pangasinan and Leyte), with principal focus on Pangasinan. A third test of RATES and CET would take place in the Maguindanao district of the island of Mindanao, with separate evaluation by Notre Dame University.

Lessons would be broadcast in Pangasinan and Leyte, as the project moved to full test of the radio model. Cassette tapes would be used to deliver radio materials in Maguindanao and in a number of additional experimental studies outside of the principal pilot test.

As the pilot test of radio-assisted instruction went forward, the Bank and the government of the Philippines would be planning together for the PRODED education project. The timetables for this planning were emerging. EDPITAF would develop and submit a long-term educational development plan to the Bank by October, 1979. A Bank team would conduct a pre-appraisal mission in November of the same year, followed by a full appraisal mission a year later — in October, 1980. Negotiations on the final shape of the project would follow in the Spring of 1981, with decisions reached before summer.

The timing of the pilot project fit this schedule. Preliminary summative evaluation data would be available in June, 1980, three months after the end of the pilot test and some four months prior to the full appraisal mission. The final summative evaluation report would be ready by December, 1980, prior to loan negotiations. The pilot project would end in July 1981, and the loan would officially close in December of the same year. The timing was tight, but the decision schedule which drove the radio pilot project could be met.

IMPLEMENTATION SUMMARY

In the pilot year RATES materials were used in 120 Grade IV classes (4,800 pupils) and 30 Grade V classes (1,100 pupils). CET reached 2300 teachers for the Educational Trends course in Pangasinan and Leyte; and 2500 for the Pilipino course in the same sites. An additional 900 teachers enrolled in these courses in Maguindanao. Numerical goals for the reach of the pilot project were met.

The materials used in the pilot test reflected the lessons learned from formative evaluation during the formative year; project operations were also improved by the experiences of the first year.

IMC successfully met demanding schedules for data collection and analysis, and produced both preliminary and final reports on schedule.
Thus, in terms of output targets and schedules, the pilot project was successfully implemented. This was a remarkable accomplishment. On the ECO side, an effective broadcast educational development organization had been created in eighteen months. The IMC had managed a large and complex field experimental evaluation in two widely-separated sites, with large samples and a very large volume of data.

THE PILOT TEST BEGINS

ECO and IMC staff were still struggling with a number of implementation problems as the pilot test began with the school year in June, 1979. The ECO studio could not be used for recording, forcing continued reliance on contract facilities. There were still no project vehicles. Fellowships for staff training could not be used because of government restrictions. Procurement of cassette players for the schools was delayed.

Administrative problems began to appear in the financial relationships between ECO and its principal subcontractors, IMC and UPLB. IMC policies for expenditures did not always correspond to those of EDPITAF, causing some friction and need for adjustments. A similar situation existed between EDPITAF and the CTRE project, where problems were occurring in the flow of funding through the university to CTRE.

Work went forward despite these difficulties. ECO began to reflect on the future of radio education, commissioning a policy study of options for radio education in the context of overall directions for development of Philippine education. 1/

Significant implementation problems were encountered in the broadcasting component of the project. Radio signals were weak, resulting in erratic reception and poor quality. This was compounded by frequent power failures which cut off broadcasting completely.

The Mass Media Specialist visited the Philippines in November, 1979, for a review of progress. His report 2/ noted problems with the studio and cassette players, as well as the relatively low level of use of technical assistance resources. Overall, the project had been "...implemented in general as scheduled."

He also reported that IMC would provide data from the formative year evaluation in early December, 1979. This would help ECO look to the future: "According to [the data] ECO will make follow-up plans of this pilot project. ECO expects these plans to lead to a possible educational media or material project, which will be an independent or supporting item of an education project (e.g. the program for decentralized educational development). ECO will modify these plans taking into account ... summative evaluation's succeeding data, ... which (probably coming in March 1980) will refer to the combined use of textbooks and radio."
The Evaluation Expert returned to the Philippines in late November and early December as a member of the Bank's preappraisal team for the WB VII project. During this visit he found time to review progress in the pilot project, particularly with respect to the summative evaluation. He had maintained contact with the Evaluation Director at IMC by mail and telephone in the period leading up to the summative evaluation.

The pre-appraisal mission was led by an economist with the East Asia Education Projects Department. This Appraisal Team Leader was aware of the radio pilot project and the schedule for completion of evaluation from Bank reports and briefings from the Mass Media Specialist. 3/

EDPITAF had prepared a comprehensive proposal for the Seventh Project in October, 1979, prior to the arrival of the pre-appraisal mission. This proposal became the basis for continuing discussions between the Bank and the government over the next eighteen months. Because of the relatively early stage of the pilot project, radio education did not figure prominently in the government's proposal at this time.

At this stage in the process of planning for the large education sector loan which would be known as PRODED little formal attention was given to the radio project. The critical point would, as planned, come a year later with the completion of the pilot year evaluation and the arrival in the Philippines of the formal Bank Appraisal Mission.

TOWARDS DECISIONS

The first six-months of 1980 promised to be significant for the pilot project. The evaluation would be completed and preliminary data would be available to assist the government in reaching decisions on the shape of educational development for the next five years.

IMC completed data collection at the end of the school year in March, and began data analysis. The Evaluation Expert continued to provide technical advice on the evaluation. In April he wrote to the IMC Evaluation Director on a number of points growing out of his visit the previous December. He noted the need for more test items to assess listening and speaking skills, a technical need for larger control groups, and the need for the cost analysis to deal more explicitly with the economies of scale and marginal costs associated with project expansion. 4/ Given that data collection for the pilot test was completed, it is clear that he was looking forward to further evaluation in the third project year (1980-81).

Difficulties in financial relationships between ECO and IMC increased during this period. IMC had difficulty in complying with government financial regulations; as a result, EDPITAF had to cut off cash flow to IMC for more than a year. 5/ The continuing discussions between the two agencies on this problem occupied considerable attention, and together with the lack of funds for IMC kept the two at some distance from each other. Technical and administrative coordination continued, but probably at a lower level of effectiveness than might have been achieved absent the funding problem.
IC submitted preliminary summative evaluation findings to ECO in mid-June, followed by a joint IMC/ECO review of the data. On July 11 IMC sent ECO a letter summarizing the gain score data for the various treatments. The interpretive language accompanying these data did not change materially from that point forward into the final IMC report on the pilot year. Thus the data reported, and their interpretation, can be taken as conclusive at that point given the strengths or weaknesses of the evaluation design and data analysis methods.

**Summative Evaluation Findings**

The data presented are shown in Tables 1 and 2 below. They are clearly equivocal. While they show substantial learning gains for various treatments, and while certain treatments scored better than the pre-test/post-test control group, the data raise important questions.

For example, RATES alone led to a gain score of 11.92, out performing CET alone at 9.99. The two treatments together led to a higher gain score, 13.13. This makes sense given the design of the intervention. However, adding textbooks on a 2:1 ratio to the RATES/CET combination resulted in a lower gain score (9.33) than either RATES alone, CET alone, or RATES with CET. This would seem to indicate that the textbooks lowered the effectiveness of the radio components. The treatments combining CET with a 2:1 textbook ratio scored slightly higher than CET with a 1:1 textbook ratio. Although the difference in scores does not appear to be statistically significant, one might reasonably expect a treatment where all students had their own textbook to yield higher scores than one where two students shared a text.

Despite these anomalies, the data did indicate that those treatments with a radio component in any configuration scored higher than either configuration of textbooks alone.

This would have been more encouraging had the post-test only control group not scored higher than two of the radio treatments and both of the textbook interventions. A partial explanation for this phenomenon might have been found in control group "contamination" caused by teachers in the control areas working extra hard to "beat" the experiment. This had been seen during the formative year. But the modifications to the pilot year design, which included the addition of a post-test only control group outside the experimental area, apparently did not control for or measure this factor.
Table 1: Mean Total Pretest and Posttest Scores of Grade IV Pupils in Pilipino, by Treatment and Combination, Pangasinan

<table>
<thead>
<tr>
<th>Treatment Combination</th>
<th>Pretest</th>
<th>Posttest</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. RATES</td>
<td>27.59</td>
<td>39.51</td>
<td>11.92</td>
</tr>
<tr>
<td>2. CET</td>
<td>29.96</td>
<td>39.95</td>
<td>9.99</td>
</tr>
<tr>
<td>3. 1:1 Student-textbook Ratio</td>
<td>31.07</td>
<td>39.76</td>
<td>8.69</td>
</tr>
<tr>
<td>4. 2:1 Student-textbook Ratio</td>
<td>27.97</td>
<td>36.49</td>
<td>8.52</td>
</tr>
<tr>
<td>5. RATES and CET</td>
<td>27.11</td>
<td>40.24</td>
<td>13.13</td>
</tr>
<tr>
<td>6. RATES and 2:1 Student-Textbook Ratio</td>
<td>30.88</td>
<td>43.02</td>
<td>12.14</td>
</tr>
<tr>
<td>7. RATES, CET and 2:1 Student-Textbook Ratio</td>
<td>25.11</td>
<td>34.44</td>
<td>9.33</td>
</tr>
<tr>
<td>8. CET and 1:1 Student-Textbook Ratio</td>
<td>29.67</td>
<td>41.31</td>
<td>11.64</td>
</tr>
<tr>
<td>9. CET and 2:1 Student-Textbook Ratio</td>
<td>28.39</td>
<td>40.24</td>
<td>11.85</td>
</tr>
<tr>
<td>10. Pretest-Posttest Control Group</td>
<td>27.42</td>
<td>37.56</td>
<td>10.14</td>
</tr>
<tr>
<td>11. Posttest Only Control Group</td>
<td>-</td>
<td>41.42</td>
<td>-</td>
</tr>
</tbody>
</table>
### Table 2: Ranked Gain Scores by Treatment Combination

<table>
<thead>
<tr>
<th>Treatment Combination</th>
<th>Gain Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. RATES and CET</td>
<td>13.13</td>
</tr>
<tr>
<td>2. RATES and 2:1 Student-Textbook Ratio</td>
<td>12.14</td>
</tr>
<tr>
<td>3. RATES</td>
<td>11.92</td>
</tr>
<tr>
<td>4. CET and 2:1 Student-Textbook Ratio</td>
<td>11.85</td>
</tr>
<tr>
<td>5. CET and 1:1 Student-Textbook Ratio</td>
<td>11.64</td>
</tr>
<tr>
<td>6. CET</td>
<td>9.99</td>
</tr>
<tr>
<td>7. RATES, CET and 2:1 Student-Textbook Ratio</td>
<td>9.33</td>
</tr>
<tr>
<td>8. 1:1 Student-Textbook Ratio</td>
<td>8.69</td>
</tr>
<tr>
<td>9. 2:1 Student-Textbook Ratio</td>
<td>8.52</td>
</tr>
</tbody>
</table>
IMC staff also pointed out that the highest mean post test score (43.02) indicated poor overall pupil performance: the average student in this treatment answered only somewhat more than 50% of the test items correctly.

The preliminary cost effectiveness data were also interesting. Using cost per point of gain score achieved, the treatments using only textbooks were most cost-effective at P 0.20 for the 2:1 ratio and P 0.31 for the 1:1 ratio. CET and a 1:1 textbook ratio cost P 0.74 per pupil point gained; RATES and a 2:1 textbook ratio cost P 13.33.

Thus while the data indicated that radio made a small difference in learning gains, textbooks were much more cost-effective. Moreover, the differences between the radio treatments and the control group were not large. And even though the control group outscored both textbook groups, the data fell short of providing strong support for the effectiveness of radio.

Both the IMC staff and the Evaluation Expert (in later correspondence 7/) believed that more thorough analysis of the data to control for a number of key student characteristics would help explain some of the anomalies in the data. Continued evaluation in the third year of the project would be important to full assessment of the effects of radio. It was also felt that impact would increase after two years of programming: one year might have been too short a time for the various treatments to have major effects. Clearly problems with broadcasting weakened the pilot; equally clearly, broadcasting reliability was a key system element to be tested.

The design of the summative evaluation created a heavy logistical burden in test administration and data collection in two geographically distant sites. IMC was hard pressed to carry out the evaluation as designed given lack of vehicles and heavy field activity demands. One result of this is an apparent lack of organized qualitative information about RATES and CET in the field. Although the ECO Project Completion Report in 1982 would provide useful summary information about project operations, these data were not available at the time of the preliminary summative evaluation in an integrated form. It was difficult to fully interpret the learning gain data without this qualitative background.

Impact of the Findings

Given the passage of time, and subsequent events, it is difficult to state with certainty what the impact of these findings was at that point during the project. The Mass Media Specialist completed a supervision mission towards the end of June. He reported that the IMC data were "informative." His report also noted that the summative evaluation was not presenting "decisive" data. 8/ It seems clear, at least, that the data did not lead the Bank staff most closely associated with the project to strongly supportive conclusions.

In the Philippines, ECO and IMC staff were more hopeful. The first two years of the project had led to significant accomplishments and a strong project team. Third year implementation could yet show strong impact. IMC agreed to conduct a follow-up summative evaluation of impact during the third
project year (1980-81). This would include assessment of impact on students with two and three years of cumulative exposure to radio education, and there would be the opportunity to strengthen both the evaluation design and data analysis methods.

APPRAISAL AND DECISION

Meanwhile, as ECO and IMC moved into third year implementation, planning for PRODED continued.

The full Bank appraisal team arrived in September and stayed into October, 1980. Working from the EDPITAF proposal of the prior year, and the work of the Bank pre-appraisal team, the mission collaborated with the Philippine government in the development of a five-year program to strengthen elementary education. This program would be the first half of the government's projected ten-year effort.

The mission was led by the same Appraisal Team Leader of the year before, and included other Bank staff from the East Asia Education Projects Division. The team also included a consultant on textbook publishing. There was no Bank official with prior direct experience with the radio pilot project. The Appraisal Team Leader had been briefed before departure on the status of the radio pilot.

The Mass Media Specialist was present in the Philippines on a supervision mission for the radio project during part of the period in which the appraisal team carried out its work. Not formally part of the team, however, he played no major role in its activity. He did seek to help ECO generate support for the radio pilot within the Philippine government, though with little success. 9/

This Bank mission was the event towards which the radio pilot project had, from the beginning, been directed. During the visit the government and the Bank would decide on the shape of investment in education for the next five years.

The draft appraisal report, and the subsequent final version, outlined a large and comprehensive program of investments to strengthen the elementary education sector. 10/ The strategy rested on policy changes, management improvement and a series of actions to improve sector performance. There was an emphasis on improving equity through investments in disadvantaged regions. Bank funds would be used to finance curriculum development; development and production of instructional materials; facilities and equipment; the monitoring of teacher supply; staff development; and sector evaluation. A key feature would be the transformation of the existing Textbook Agency into an Instructional Materials Development Corporation.
The radio project was noted as an ongoing Bank-supported effort. However, the appraisal report made no substantive reference to the pilot project and radio was not incorporated into the PRODED project.

The Pilot Project Director recalls being given a week's notice from EDPITAF leadership to develop a proposal for use of radio prior to the Bank team's arrival. This proposal outlined region-based use of radio and tape-assisted instruction to strengthen language arts instruction in Filipino and English, a strategy consistent with the thrust of PRODED towards assistance to educationally disadvantaged regions. The report was given to concept with the Appraisal Team Leader. She reports that the team was heavily focused on textbook production with strong leadership from the textbook consultant; the radio component appeared to complicate the strategy and thus did not become an integral part of the appraisal report.

A major component of the proposed strategy was the creation of an Instructional Materials Development Corporation (IMDC). This entity would be responsible for developing, producing and disseminating learning materials. Although the report emphasized print materials, ECO and EDPITAF staff saw the corporation as a possible structure through which radio education might eventually be expanded. While the emerging PRODED effort did not explicitly include broadcast education, it did not exclude it either.

**Position of the Philippine Government**

Planning for a Bank financed project is a collaborative process. EDPITAF and the Ministry of Education and Culture had a great deal to say about what went into PRODED. Clearly, government support for the radio pilot project was weak at this point.

Leadership had changed during the preceding two years. The individual who was EDPITAF Director when the pilot project was initiated, and a strong supporter, had left the agency in 1979 and had passed away in 1980. A new EDPITAF Director has been appointed. He supported the project, but he did not have the same history of close involvement in its development.

Moreover, EDPITAF was approaching the last year of its mandated existence. The future of the agency was uncertain. Much time and attention was necessarily being devoted to the larger issue of EDPITAF's future. The Minister of Education was to retire at the end of 1980. With his successor's attitudes towards radio education unknown, the leadership situation was further clouded.

Thus the pilot project had no strong senior advocate within Philippine leadership circles. With somewhat equivocal preliminary summative evaluation findings (the final summative evaluation report would be submitted, as scheduled, in December), a strong push for radio education did not appear possible.
Position of the Bank

Nor did the Bank team include an advocate for radio education. The team members had no prior association with the pilot project; none of them were expert in broadcast education. The team did have a strong advocate of textbooks, whose influence is said to have been important in the joint government/Bank decision to emphasize textbooks. 13/

Given the situation and the key actors in the appraisal process, it is not surprising that radio education did not become a formal part of PRODED at that time.

The Mass Media Specialist and the Evaluation Expert maintained their continuing interest in the project during this period. At the time that the appraisal mission was in the Philippines, ECO received a cable from the Bank passing on the outline of the proposed book about the project. 14/ The Mass Media Specialist was scheduled for a supervision mission in February, 1981. The formal summative evaluation report was yet to come out, and the Evaluation Expert was in correspondence with IMC in September on refinements in the third year evaluation design. But not a part of the project appraisal process, and working out of different divisions, their direct influence in the decision-making process was minimal.

CONTINUING EFFORTS FOR RADIO EDUCATION

Meanwhile, ECO continued with implementation of the third year of RATES and CET. Dissemination switched from broadcasting to classroom use of cassette tape, partly in recognition of difficulties encountered in broadcasting and partly as a trial of an alternative delivery system for the future.

IMC conducted the third year evaluation under continued contractual relationship with EDPITAF/ECO. This evaluation reflected changes growing out of the pilot year experience. The number of treatments was reduced, simplifying the design. Data analysis was more sophisticated, including a number of multivariate analysis methods. Qualitative data were obtained through unscheduled visits to project classrooms. A post-test only control group in a different area was added to control for contamination due to extra effort by control group teachers in the experimental area, a feature that had also been added to the pilot year evaluation design.

The final report for the third year evaluation indicates that two continuing problems in the design were not fully addressed. The cost analyses still did not include data on regular classroom instruction. There is no indication that the learning gain tests were revised to include more items directly measuring the impact of radio.
The third year evaluation results were similar to the findings of the pilot year. RATES students scored slightly better than students in other treatment or control groups. 15/

The financial problems between ECO and IMC were resolved through an audit of IMC records, with final adjustments being incorporated into the final year's contract. This contract, unlike the previous ones, was funded with Bank funds as opposed to Philippine government counterpart resources.

The project officially ended in July, 1981. The loan closed for further expenditures in December of the same year. IMC submitted the first draft of its evaluation report in January, 1982. ECO began drafting its Project Completion report at about the same time.

These are the major events and dates in the project record during the final year. Planning for PRODED went ahead, with the final Bank appraisal report being completed in April, 1981. Negotiations were completed for the project to begin in 1982. Radio was not formally included in the final design of the project.

The Search for Support in the Philippines

The effort to see the pilot project forward to a useful conclusion did not stop with the Bank appraisal mission. Though disappointed by the outcomes at that stage, ECO and EDPITAF continued to seek support for the project and to evolve a pattern of follow-up action to utilize project outputs. The objective of establishing radio as an important part of elementary education had not yet been achieved, but the staff's belief in the project had not lessened. The materials (in broadcast and cassette format) were having an impact, however small. Moreover, much had been accomplished towards the Philippine objectives of service and institution-building. It was important, at least, to capitalize on these accomplishments.

A new Minister of Education was appointed in January, 1981. He came to the post from the presidency of the University of the Philippines. In that capacity he had been aware, administratively, of the radio project through IMC's role in the evaluation.

ECO sent copies of the IMC final evaluation report to the new Minister, and began an informal campaign to develop support for the project. 16/ There was still time to have radio included in PRODED. The proposed Instructional Materials Development Corporation could, for example, develop and provide radio materials as well as textbooks.

In line with the regional development strategy of the Ministry EDPITAF requested proposals for continuation of radio education from the regions that had been involved in the pilot program. In a memo to Regional Directors dated March 20, 1981, the EDPITAF Director noted: "As you have been informed by the EDPITAF/ECO staff, the radio education project will officially end in July, 1981. However, we shall gladly assist you in preparing your regional plans for the continued implementation of RATES and CET beyond 1981 and in identifying the resources you may need. We shall also look into possible provision of some of these identified resources." 17/
Regional support for RATES and CET was strong, and all three regions responded with proposals in March. While they differed in detail, they all proposed continuation and expansion of use of RATES and CET materials on a tape-assisted basis. The projects would be carried out within the context of PRODED. These proposals provided the basis for continued use of the materials developed, and would serve as the framework for transition from the ECO pilot project to the future.

Efforts to develop top-level support for radio education faltered. The Project Advisory Group (PAG), which might have served as a bridge between EDPITAF/ECO and the Ministry of Education, began to wind down as the project neared completion. The PAG was in part held in place through honoraria payments; with project funding ending, this basis disappeared.

EDPITAF itself was in a critical transitional period during which the government was studying alternatives for the future of the agency. Direct management of EDPITAF activities from the Ministry gradually increased during the year, culminating with the official appointment of a Vice-minister of Education to lead EDPITAF in December, 1981. The EDPITAF Director recalls the uncertainty of communication and decision lines during this period. He noted that, as the radio project neared completion, he attempted to conserve the large balance of unspent Bank funds in the project to support "technician education," hoping to establish a basis for continuation in a different framework. This strategy did not succeed; the Bank eventually cancelled the outstanding amount of the loan. 18/

The new Minister was not supportive of educational radio. ECO attempts to secure support for the project met with little success. Planning proceeded without an educational radio component.

The Summative Evaluation Reviewed

The pilot project ended in July, 1981. An additional six months were available in which Bank funds could be used to close out project activities. ECO staff used this period to transfer project materials and resources to the three cooperating regions to support their plans for continuation. IMC completed data analysis for the third year evaluation, preparing their report for submission.

ECO retained two consultants from the US Northwest Regional Educational Laboratory to advise ECO on guidelines for continuing media-assisted education. In the course of their work the two consultants commented favorably on the instructional designs used, on the quality of materials and on formative evaluation efforts.

They also reviewed the pilot year summative evaluation report in search of lessons which would help guide future activities. Time prevented a comprehensive examination of the entire five volume evaluation report. However, their findings are of interest, shedding some light on issues regarding the design of the evaluation and its utility for policy decisions. 19/
Their conclusions are reflected in the following excerpts from their report:

"We conclude that little information is provided in the summative evaluation that is useful in making decisions about the further use, modification or development of ECO materials. Because of the large effort that the evaluation represents, we would have welcomed any opportunity to have reached a different conclusion."

"At one level it appears that the evaluation has a clear purpose that is served by a large scale experimental study. Upon further examination, the clarity of planning becomes blurred. First there is no indication of the audience for the evaluation or the information needs of the audience. This results in a lack of a rationale for conducting the evaluation. Second, little thought has been given to the methods of Filipino instruction which could be used in comparison with the RATES program. Only control groups and combinations of RATES with other ECO programs were used. Thus, a tape-assisted approach (currently being used in many locations) and a textbook approach were overlooked as possible 'competitors' to the RATES program. In short, realistic alternatives, one of which is being extensively adopted, were ignored, and no information about the costs or uses of alternatives were provided."

"To summarize, the evaluation is narrowly based, and there is little indication about the questions that were intended to be answered by the evaluation... this lack of clarity is reflected in an analysis plan that seems to lack any direction."

"Because of the large investment in the experimental part of the study, scant attention was given to other important elements, such as qualitative information. Cost data were included in the evaluation, but the cost analysis did not provide information helpful to those who would be interested in adopting RATES."

"... the evaluation gives us no information about who can and who cannot profitably adopt RATES and under what conditions. Neither do we receive information about how RATES differs from what typically occurs in the classroom."

"According to the evaluation report and staff reports, clear contamination of control groups occurred. The control teachers apparently were aware that they were part of a study, and they responded by trying to "beat" the treatment. To some extent, the control group became an alternative treatment comprising more motivated teachers. This occurrence demonstrates the lack of robustness of complex experimental
designs .... It also demonstrates the need for evaluations to avoid relying solely on experimental studies and to adopt a flexible approach."

"In comparison to the effort apparently required to develop the sampling plan, the thought and effort devoted to the development of a criterion measure seems minimal.... Our judgement is that the test is weighted so greatly in the direction of reading items that it does not reflect the heavy listening component of RATES, making it insufficiently content valid to be used as a sole criterion measure."

"Poor data analysis is a particularly important flaw in the evaluation. The analysis approach is marked by so much fragmentation and lack of direction that is impossible to characterize simply."

In general, existing reports of the summative evaluation are not as useful as they need to be. The primary difficulty is that they reflect the lack of direction of the analysis and the absence of a clear evaluation purpose that were noted earlier."

"The Executive Summary, as now written, is excessively long and lacks focus."

These are harsh criticisms. As with all limited ex post facto reviews they do not — and could not — take into account the many contextual variables that contributed to the design and execution of the summative evaluation. Some of the observations seem in error: for example, textbooks as a competitor to radio were assessed. However, they do provide insight into the relative effectiveness of the summative evaluation reports in generating support for the pilot project.

The criticisms echo issues regarding the evaluation design raised as early as the first year of the project, issues which surfaced periodically.

The Role of the Bank

Bank staff completed three supervision mission during this period, in November 1979, in June 1980, and in February 1981.*

Supervision Missions

The Mass Media Specialist came alone on each mission. Although the Evaluation Expert visited the Philippines in connection with the pre-appraisal mission in late 1979, he was unable to participate directly in the Philippines during most of this critical period. He did continue to provide technical support to IMC by mail and telephone.

* A final review mission, which took place in November, 1981 at the very end of the project will be reviewed in the following chapter.
The Mass Media Specialist's supervision reports provide a useful review of the progress being made and problems being encountered. His report in June, 1980, came at the point in the pilot project when preliminary summative evaluation data became available. The report noted the implementation problems being encountered, but indicated that the project was being carried out "without major difficulty,"\(^{20}\) The effect of the project's uncertain future on staff morale and turnover was also reported.

The preliminary evaluation data are reported. Interestingly, Bank files show that an early draft of the report said that the data were "encouraging." This was later edited to read "informative." Given the equivocal nature of the findings, the latter adjective may have been more appropriate, but the editing indicates that a shift in Bank staff perceptions of the project may have been taking place. Competition from teachers in control groups was reported, and it was noted that the Evaluation Expert was "reviewing" the statistical analysis methods of the report.

It was also reported that EDPITAF would be soliciting plans from regional offices for project follow-up, "...but EDPITAF will not attempt immediate nation-wide application of CET nor RATES approach. The selected schemes would be considered for possible inclusion in the proposed Education VII. Perhaps about five regions would be selected by EDPITAF for following up CET and/or RATES approach. The scheme would be part of ... (PRODED)."

This report also notes that EDPITAF saw ECO's future role as a production center for core materials to be distributed to regions, perhaps in conjunction with the proposed Instructional Materials Development Corporation.

When he returned in February of 1981, the Mass Media Specialist found that the Ministry of Education and Culture was still working on follow-up plans. "Neither EDPITAF nor MEC has yet mapped out concrete schemes of following up the outcome of the summative evaluation, but the Textbook Board secretariat has agreed, in principle, with EDPITAF's plan of including in the coming seventh education project a unit, under the control of the proposed Instructional Materials Corporation, that would be responsible for distributing taped lessons as enrichment materials for Pilipino and English for grades one to six. ECO wishes that most of ECO's present resources, especially for educational materials production, could be employed by this new unit."\(^{21}\)

It will be recalled that this is the essence of the plan presented by ECO to the Bank appraisal mission the previous October. EDPITAF requested and received regional plans for continuation immediately following this February visit.

Finally, the February report provided an assessment of the pilot project as a whole: "The S-8 project, though the amount of the loan is US$2 million, is unique in that the whole project is meant for a summative evaluation of pilot schemes. The evaluation data are supposed to be properly assessed and followed up, if it is worthy of support, by the Government. It will be of no small concern to us for formulating education projects to see that this evaluation is well controlled and fully utilized, and for this purpose, it may be desirable to send an implementation review mission in late 1981."
Taken together, these supervision reports reflect good understanding of developments in the Philippines during the period. In consequence, perhaps, the tenor is somewhat less optimistic than earlier reports. Clearly the summative evaluation data had not demonstrated a significantly successful pilot; equally clearly, the Government was not moving strongly to support radio as a key strategy for educational development.

**Use of Bank Resources**

The rate at which ECO was using loan resources became a concern relatively early in the project. By July, 1980, the radio project was on a list of "underimplemented" projects noted in a letter from the Bank East Asia Education Projects Department Director to the EDPITAF Director. The project would, in the end, be underspent by a million dollars, $850,000 of which would be Bank loan funds.

Inability to procure vehicles was part of this, of course. In addition, funds remaining from the pre-investment study had been used by ECO to finance early technical assistance, reducing the draw on TA funds in the S-8 PH loan. It had been difficult to use fellowship funds because of Philippine government restrictions on study abroad for contract personnel.

A major shift in the planned use of Bank resources came in the technical assistance area. Some nine person-years of technical assistance were budgeted; only 16 person-months were used, the bulk by ECO in the development of management, production and formative evaluation procedures. IMC used only about one person month of external technical assistance during the life of the project.

Both ECO and IMC felt that good local technical assistance was available when needed to supplement staff resources. Early in the project, IMC signalled a preference not only to use Philippine experts, but to use funds as much as possible for training. This was consistent with Philippine objectives to use the project for institution-building.

In addition, IMC had collaborated closely with the Evaluation Expert in the technical aspects of developing the summative evaluation design. This significant level of technical cooperation did not show up in Bank accounting.

**Continuity in the Bank**

The degree to which the pilot project was linked to Bank operations during this period was undoubtedly important. The earlier team support clearly weakened during the latter part of the project. The Evaluation Expert changed posts during the second year of the project. New duties limited his availability for supervision missions, and in fact he visited the Philippines only once during the final eighteen months of the project — and then as part of the pre-appraisal mission for the seventh loan.
His participation in the pre-appraisal mission was important, however, as it provided a direct link between the pilot project and Bank operations for the education loan. This direct link was absent a year later when the appraisal mission was in the Philippines, reaching final decisions with the Philippine government on the shape of PRODED.
Projects do not end abruptly. There is always a period during which the momentum engendered by project activity is gradually dispersed. It is during this period that the opportunity for organized learning from project experience is at its strongest.

For "successful" pilot projects this is the time when lessons learned are applied to implementation, usually on a larger scale, of some version of the program that was tested. For less successful efforts, it is a period of uncertainty, a time during which the ambiguity of outcome and process that characterized the pilot effort roll forward, complicating efforts to learn systematically from what was attempted — and what was accomplished.

The Bank loan for the radio pilot project formally closed in May, 1982. In the five months after the project completion date IMC completed its analysis and report on the second year evaluation. ECO transferred its materials and much of its equipment to the three regional education offices which would continue to use RATES and CET on a tape-assisted basis. The ECO staff dispersed, many going to the Development Academy of the Philippines, where they formed the core staff for the Educational Reorientation Project (ERP). A significant in-service teacher training component of PRODED, ERP became a principal repository of the expertise in media-assisted education developed by ECO. The three regional projects, plus ERP, constitute an important level of institutionalization of the resources developed through the pilot project.

EDPITAF continued its evolutionary journey towards recreation as the Project Development Implementation Management Office (PDIMO), a journey which was officially completed with the formal establishment of PDIMO in October, 1983. PRODED, the government educational development effort, went forward. The Instructional Materials Development Corporation (IMDC), a centerpiece of the project, did not become a reality because of a government ban on the formation of new public corporations. As of January, 1984 the IMDC had yet to be established. 1/

LEARNING FROM THE PILOT PROJECT

In February 1981, the Mass Media Specialist had indicated the need for a Bank mission to review implementation of the pilot project at the end of the year. 2/ This review was conducted in November by a team led by the "official" East Asia Education Projects Division director of the pilot. Included on the team were a consultant and an Administrative Assistant in the East Asia Education Projects Division.
In the Philippines for two weeks, the team was responsible for the review of the large textbook project, the $100 million PRODED project, and the radio pilot. Given the number of tasks, and the relative size of the different projects, it is understandable that the team spent only one day on the educational radio pilot project. They met with no one from IMC; from ECO they met only with the administrative officer.

The report of this visit recognizes that ECO had distributed tapes and dubbing machines to the pilot regions to enable them to "... replicate the tapes and thus spread the impact thereof to additional schools in their region."

With respect to the impact of the pilot project, the report noted:

3/

"The Loan Agreement ... indicated that there should be a future education project based on the...pilot studies, but the previous supervision report indicated that no follow-up schemes had been mapped out by EDPITAF or MEC, and MEC has made no 1982 budget provision for any follow up. The project therefore remains as a pilot study only."

The report further noted that ECO staff had shrunk considerably, raising the possibility of "...further fragmentation of this effective and efficient unit."

Inevitably, this very brief review of the project missed the important developments of the nine months since the Mass Media Specialist had visited in February. Proposals for follow up efforts had indeed been developed. Three proposals had been submitted by the pilot regions, and ECO had previously developed a national proposal for review by the supervision mission.

That the radio project review was added on to supervision for the two larger projects indicates rather clearly that the East Asia Education Projects Department was not very interested in the pilot by this time.

The Bank Project Completion Report

The Bank arranged for the consultant on the November 1981 supervision team to return to the Philippines in June, 1982, to prepare a completion report on the pilot project as well as to provide supervision to PRODED. Approximately twelve working days in the Philippines were made available by the Bank for both tasks. The Pilot Project Director was asked to hold submission of the ECO Project Completion Report pending this visit. The consultant reports that the ECO PCR was not ready at the time of his visit, and that he helped work through the financial data. 4/ The final IMC summative evaluation report (including the evaluation of the second year programming) was available when the consultant arrived.
The consultant's report provides a general assessment of the pilot project. Given the time available, not much more was possible.

The report indicates that the results of the pilot project are "clouded" because of a shift from broadcasting to tape-assisted instruction. This finding stemmed from the extensive use of cassettes in the schools at the time of the review. The switch had indeed occurred, but only after the pilot year was completed. This point was clarified by ECO staff in a later memorandum to the Bank. The project held true to the design requirement of broadcasting for the critical first year of summative evaluation. While there were problems with the cost analysis, they were not caused by the shift to tapes after the pilot year. The cost data in the pilot year evaluation were based on a broadcasting model.

The Bank asked the consultant to follow a standard heading outline in completing the PCR. In combination with the short period of time available, this led to omission of a number of elements of the project which were significant because of its experimental, pilot nature. There is no discussion of the decision making and research utilization aspects of the project. The outcome data for the field experiment are reported briefly, without interpretation or discussion. The consultant reports that Bank staff agreed to provide this interpretive discussion. Somewhat more attention is given to the cost data for the various treatments, with a cautionary note regarding contamination by the "switch" from broadcasting to tapes.

At the time that the consultant visited the Philippines, the Evaluation Director was recovering from a serious illness and was not available to discussions. The consultant did not meet with any of the IMC evaluation staff.

There is no mention of the review of the summative evaluation design conducted earlier by the two consultants to ECO; indeed, the design is described but not evaluated in any way.

The draft PCR took somewhat more than twelve months to receive final editing and clearances in the Bank (the approved report is dated August, 1983).

Despite these difficulties, the project completion report reaches some conclusions which are consistent with the story of the pilot project as recounted in the present study:

"One lesson learned from the project is that highly innovative projects need close supervision and monitoring. The evaluation design should have been reviewed at various stages of implementation to ensure that necessary adjustments were made."

"Experience gained under this project suggests that the Bank is not the ideal agency to carry out small-scale, innovative projects. Other institutions with greater technical assistance resources probably have an advantage in this kind of activity."
It was conclusions of this nature which led the Bank to commission a full case study of the pilot project and its evaluation.

The conclusions are understandable, but they do not adequately reflect the richness of the pilot effort, nor the many variables which contributed to the shape of the project and its impact on decision-making.

Perhaps the most significant aspect of the Bank Project Completion Report is the scant time and resources devoted to the review at that time. That the pilot project had no advocates, and had lost the attention of the Bank, seems clear.

The Project Disbands

With decisions on the future of the project long-since taken, ECO began to disband. Key staff joined ERP, taking with them the CET teacher training materials and a high level of expertise in the design and production of media materials. Regional education offices moved ahead to use RATES and CET materials. Limited followup from EDPITAF indicated that the materials were widely and effectively used, though systematic review was not done. 9/

IMC completed the final evaluation report, covering the second year of RATES and CET.

Direct efforts to have RATES or CET included in PRODED continued, but at a very low level. The new Minister of Education had not changed his mind about radio education. There was some hope that, when established, the Instructional Materials Development Corporation would provide a mechanism through which radio materials might be developed for wider use.

A year and a half later, in January 1984, yet another ministerial change occurred. This new Minister was thought to be more amenable to use of radio: he had been a supporter of the pilot project before it began. EDPITAF leadership spoke hopefully about a second chance. 10/

SUMMARY DISCUSSION

A review of the period beginning with the pilot year and ending with the consultant’s visit in June, 1982 leads to the conclusion that, intentionally or not, the Bank gradually abandoned the pilot project after an initial period of great enthusiasm and significant support.

The continuity represented by key Bank staff working closely with project managers and evaluators began to erode as other duties increasingly prevented the Evaluation Expert from direct participation. The Mass Media Specialist represented a supportive and continuing presence, but not an evaluation expert and isolated from decision-makers in the East Asia Educational Projects Department, he could not serve as an effective link for the project.
Towards the middle and end of the project, key Bank missions to the Philippines did not involve either of the two Bank staff most closely involved with the pilot project. In the end, the final two missions were staffed by individuals with very little familiarity with the project and its history. The reports of these missions, given the limitations noted, do not support much learning from the pilot project experience.

As the project began, the Evaluation Expert and the Evaluation Director planned a significant book on the pilot project. It would document success, and serve as an important contribution to the literature on radio education. The book was a live issue as late as October, 1980, when a detailed outline of its contents were cabled from the Bank to the IMC. By the end of the project the book was an afterthought. In early 1984 a manuscript existed, but the IMC was reluctant to publish given no financial support from the Bank and little apparent interest. 11/

In the endgame period, decisions had already been made regarding the role of radio in PRODED. Radio was not included. No mechanisms existed in the Bank, or in the Philippines, to keep the project visible other than the continuing interest of key actors, notably the Pilot Project Director and, to some extent, the Evaluation Director.

The project was designed to climax with decisions following the pilot summative evaluation. The subsequent year was in place primarily to enable the project to move forward as part of the five-year educational development effort. The design, based on optimism and the expectation that the pilot would be incorporated into the educational system, made no provision for review and learning activities in the event that the pilot should fail to have the desired impact. In consequence, little was done.

The decline in Bank interest and involvement has been discussed. On the Philippine side, there was little formal interaction in the endgame period beyond that necessary to meet administrative completion requirements. The ECO Project Completion report was finished and submitted. Copies were sent on a limited basis to key individuals. There were no seminars, no panels, no inquiries. In the Philippines, as in the Bank, the project appeared to fade away quietly, with only a few shadows to mark the spot where so much energy and commitment had been expended.
Chapter Five: Patterns and Lessons

Several clear patterns emerge from the story of the Philippine radio education pilot project. These help us understand what happened and why. In turn, within the limits of case study analysis, they enable us to draw lessons from the experience which may help with similar future projects.

PATTERNS

A large-scale pilot project was established to assist with decisions on alternative ways to strengthen basic education in the Philippines. When those decisions were taken, the results of the pilot seemed to have been little considered. Why this occurred is the fundamental question to which this study is addressed. The patterns drawn from the narrative analysis should help us in reaching answers to that question.

As is often the case, there is no single — or simple — answer to the question. To the extent that an explanation exists, it is constructed from insight generated by the interaction of several dynamic dimensions of the pilot project experience. These dimensions — or patterns — interacted with each other, and each is made up of complex sets of decisions, actions and circumstances.

In sum, these patterns appear to fall within three categories: decision-making structures and processes, the design of the pilot project, both in implementation and evaluation, and interactions among key individuals and agencies engaged in the effort. No single category suffices as an explanation. Taken together, they provide at least a partial answer.

Decision-making Structures and Processes

No formal mechanisms were built into the project to ensure that outcomes would be used in reaching decisions on the shape of continuing educational development in the Philippines. When the project was planned there appears to have been an implicit assumption that the key actors and decision structures present at inception would continue through the life of the pilot, and be in place when the results became available.

This assumption, of course, turned out to be ill-founded. There were significant personnel changes in EDPITAF and the Ministry. The key Bank official in the design of the project became increasingly divorced from its progress and outcomes. The broad Philippine agency involvement that characterized the initiation of the project dissolved with implementation, in part because of the narrowed focus that the pilot represented in relation to earlier, more ambitious plans.
The single project mechanism that might have served to focus decision-making on outcomes — the Project Advisory Group — was too narrow in membership to create widespread and continuing interest in the project during its implementation, and went out of existence just at the time when it would have been most useful. The implementing agency — EDPITAF — went through complete reorganization precisely at the time when stability was needed to ensure attention to the radio project.

Project supervision in the Bank was located outside of the department charged with the larger questions of educational development in the Philippines. This distance, and the small size of the pilot, led to a situation in which the effort received relatively little attention during implementation. In the Bank, as in the Philippines, there was no recognized and persistent mechanism for integration of the pilot project into ongoing decision-making.

The IMC did not play an active role in seeking decision-oriented use of the evaluation. Such a role was not built into the project in the beginning. It may also be that the nature of the evaluation findings, as well as problems with the design, reinforced this position within the IMC.

IMC staff worked very hard to prepare an extensive executive summary of the evaluation findings. Although criticized as too long, the summary could have been used as the basis for presentations of the data. For example, the presentation format in the document facilities the making of overhead transparencies. Yet the sheer bulk of the evaluation report makes its use difficult, and quite probably contributed to its lack of impact in the decision process.

At the end, when the shape of PRODED was being determined, the lack of formal mechanisms combined with personnel changes led to a situation in which the pilot had no senior advocates able to obtain thorough consideration of project outcomes. The project had become, in effect, an orphan.

Design

The absence of explicit decision-making structures is, in retrospect, a flaw in the design of the pilot project. There are other aspects of the design of the pilot which help us understand the outcomes. These relate to the design of both the intervention and the evaluation.

Fundamental to the design of both the intervention and the evaluation was the focus of the effort. The pilot was, from the point of view of the Bank, designed for the evaluation. The purpose of the evaluation, in turn, was to answer this question: "How do CET, RATES and additional textbook availability — introduced separately or in combination — affect student performance?" Philippine leaders saw the project as providing service to education and contributing towards the development of institutional capability in broadcast education. These objectives were compatible. Yet neither approach focused on the principal question of
interest for long-term educational development in the Philippines. This might be phrased as follows: How cost-effective is radio as an alternative delivery system for rural primary education in the Philippines?

A clear answer to this question would have facilitated decision-making for PRODED. A clear and positive answer might well have led to incorporation of radio strategies; a clear and negative answer would have dispelled some of the lingering doubts as to whether or not the pilot had served a useful purpose.

But the pilot was not focused on this question. The intervention itself was designed to answer other numerous questions. Hence instead of a much simpler intervention design focused on radio as the variable, there was a complex design testing radio for direct instruction for Pilipino in the classroom, for teacher training in several subject areas, and in analysis design prevented comparison of any outcomes with standard classroom practice, further weakening the design in answering the principal question.

The large scale, multi-treatment field experiment appears to have been the model of choice for two reasons. One was an overt desire for a high level of "scientific" certainty in conclusions. Another, more implicit reason was the desire to contribute substantively to the growing world body of knowledge on radio education. Had it worked perfectly, this pilot could have been a landmark effort, advancing knowledge of educational radio on several significant dimensions. One is the combination of radio teacher training with direct classroom instruction. Another is the use of radio in combination with new textbooks. High levels of certainty in conclusion would, of course, have strengthened these contributions significantly. The book growing out of the project could have been a major contribution.

The sheer size and complexity of the project needed to carry out the design consumed the managerial and time resources of the implementers. Neither ECO nor IMC could do much beyond fulfilling (very well) the original plan, particularly given the difficulties encountered in procuring vehicles and equipment. This latter problem greatly affected the efficiency of the implementing agencies.

The IMC, as a mass communication research agency, undertook to evaluate an educational intervention. Many of the components of evaluation of this type differ from mass communication research: for example, learning test construction and classroom observation. While there is little doubt that IMC could have developed competencies in these areas, the struggle to carry out the basic plan left little room for growth of this type.

For example, it would be useful to have data on the amount of time students spent studying Pilipino in the different treatment and control groups. Time on task is a key variable in learning, regardless of medium. Data on this variable might help explain why CET and a 2:1 textbook ratio led to higher learning gains than CET and a 1:1 textbook ratio. Random classroom observations and interviews with students, teachers and parents could have provided insight in this area. Classroom observation was indeed a feature of the third year evaluation, but was not there for the decision year.
The many treatments led to test instruments which sought to measure both reading and listening skills. The inclusion of textbooks required this: the balance of items and their sensitivity to the effects of radio are questionable. The development of very good tests is a difficult and time-consuming task. It is possible, at least, that the size and complexity of the design resulted in less attention to this component than was needed.

ECO conducted a number of small independent studies during the pilot year. One, for example, assessed the impact of tape-assisted CET lessons for teacher training — and found good results. A simpler summative evaluation design could have freed human and fiscal resources for a larger number of smaller, decision-focused studies of this nature, yielding a richer base of information for decision-making.

The size and complexity of the pilot, in combination with a lack of focus on the principal question, led to a situation in which evaluation findings were equivocal. Equivocal findings, in turn, weakened the position of remaining project advocates vis a vis the decision-making process, thus compounding the absence of formal decision-making structures.

It is interesting to reflect on the outcomes of the evaluation effort had there not been serious disruptions in broadcasting during the pilot year. Teachers and students would have had more, and better quality, exposure to radio. Impact might have been higher. The summative design, whatever its weakness, would have captured this. Clearer and larger differences between radio and non-radio treatments would have lent strength to arguments for continuation and expansion of the project. There still would have been problems in comparing the cost-effectiveness of treatments with standard classroom practice, and one would still have questioned the validity of the test instruments. These issues, however, would be raised with the possibility of even greater gains in mind. Control group contamination would be seen in the same light.

Viewed from this angle, a weakness of the design lay in its optimism regarding the implementation of the pilot. Pilot projects everywhere test both the concept of the intervention and the mechanisms for its implementation. Pessimism on both counts might be a more appropriate initial stance. In this case, and perhaps in others, it would point toward designs which are simpler, freeing resources to focus more powerfully on the key conceptual and operational linkages in the intervention.

Interactions

The overall pattern of interaction among key individuals and agencies in the project is one of increasing distance over time. The pilot began with very close collaboration between the Bank, EDPITAF, IMC and other key Philippine agencies. Much of this collaboration was supported by close interpersonal contact among staff in the Bank and in the Philippines.
This pattern began to dissolve almost as implementation began. The Evaluation Expert's participation dwindled over time, leaving the Mass Media Specialist as the principal linkage. He, in turn, had little influence over the fundamental dimensions of the pilot design. With the creation of ECO, direct involvement of senior EDPITAF leadership lessened gradually, then at an accelerated pace when the original Director left the agency.

Differences between ECO and IMC over the summative evaluation design emerged within the first six months of the project, and were exacerbated by administrative difficulties. With IMC leadership working directly with the Bank's Evaluation Expert, ECO staff lost any substantive control over the evaluation. What began as a collaborative relationship became an arms-length, formalized pattern of interaction.

In addition, the changes through which EDPITAF was passing made the pattern for utilization of the evaluation unclear. Not only did the ECO Project Advisory group phase out when it might have helped in developing the formal review of the evaluation, but the general relationship between EDPITAF and the Ministry of Education was in flux. During much of 1981 the Ministry was assuming more direct control. This weakened EDPITAF's ability to bring issues forward strongly.

Close collaboration between Bank and Philippine staff was important to the initiation of the pilot project. However, it may have established a pattern in which significant modifications in the pilot design could come only with direct participation of Bank staff, particularly the Evaluation Expert. This pattern further reinforced the rigidity of the original design and ECO's lack of influence over the directions of the summative evaluation. It helps explain, for example, why the issues of collecting cost data on standard classroom practice and relevance of learning gain tests to radio teaching were never resolved.

Summary

A summary statement of the significant patterns in the project might go something like this. Hampered by a lack of clear focus on the most relevant question, an absence of formal decision-making structures, an extremely complex intervention and evaluation design, a lack of continuity of key personnel and increasing distance between key agencies, the project nonetheless succeeded in developing a large body of educational material and demonstrating, with some degree of certainty, that radio and cassette education was feasible and had some impact in the Philippines.

For the pilot project to have been judged a success there would have to have been significant differences in more than one, and perhaps all, of these patterns. Formal decision-making mechanisms alone, for example, might have helped with the problems of continuity of personnel and distance between agencies but could not have overcome the complexity of the design and lack of focus. A clearer focus might well have led to a simpler and more effective design; a simpler design would have facilitated more flexibility
and, perhaps, greater impact and less equivocal findings. Absent continuity of personnel and formal decision-making mechanisms, this still might not have resulted in the adoption of radio as a viable alternative for educational development.

LESSONS

Case studies are, by definition, limited to a specific phenomenon in a given place at a given time. The validity of the insights they generate is limited to similar phenomena under similar circumstances. The transfer of insight from a case to another situation requires an act of interpretation in which the context of the case is compared with the context of the new situation. It is for that reason that cases must be rich in detail, and that application of lessons from a case to a new situation involve comparative analysis and judgement from the reader.

Given these caveats, there are lessons in this case for the design of pilot projects. They are most immediately relevant to the World Bank and its partners in developing countries: a principal contextual variable in this case is the partnership between the Bank and developing nation agencies. The lessons should also be useful in other situations involving a funding agency with technical expertise on the one hand, and implementing agencies on the other.

Some of the lessons apply to situations involving multiple agencies; others may be more generally useful in pilot project design even when a single organization is involved. Many of the lessons are not new and may be found in various parts of the evaluation literature.

1. **Pilot projects should be clearly focused on policy decisions.**

   The lack of clear focus in the Philippine radio pilot had significant impact on the nature of the pilot design. Although the evaluation question was relevant, it was not asked in a way which centered attention on radio in comparison with alternative delivery systems. This led to a diffuse and much too complex design for the intervention and the evaluation, which in turn diffused energies and attention.

2. **Pilot projects should be no more complex than necessary.**

   A clear corollary of sharp focus is the potential, at least, for simplicity and economy in design. This can facilitate enroute flexibility to adjust to changing circumstances, maximizing the potential to learn from the pilot experience. The complexity of the design in the Philippine pilot clearly hampered flexibility of this kind.

   It is generally understood that the technical sophistication of any project should not exceed the competence of implementing agencies. The Philippine case suggests that, in fact, technical demands of pilot projects should fall well within the demonstrated capabilities of implementors.
special challenges to flexibility and management imposed by pilot projects consume significant amounts of time and energy. To the extent that technical demands strain agency capabilities as well, something will have to give — either flexibility or technical excellence.

This seems particularly true where the time available for the pilot is constrained by a rigid decision schedule. The Philippine agencies were extremely competent and skilled, undoubtedly able — with a flexible time schedule — to identify and deal with problems both of design and implementation.

3. Conceptual modelling of both the intervention and the evaluation is essential.

Much more attention was given in the Philippine pilot to evaluation design than to modelling for the intervention. A number of key assumptions in the supposed causality linking radio interventions and improvements in pupil learning were not examined. For example, it was assumed that a semester of radio lessons would so change teacher classroom performance that student learning would improve significantly, yet there was no mechanism to assess change in teacher behavior in the classroom.

4. Mechanisms for decision-making should be an integral part of project designs.

Individuals and organizations change. Pilot projects should have strong internal and external mechanisms which enable them to adjust to these changes enroute, ensuring an audience for their outcomes. Such mechanisms, moreover, facilitate the ongoing education about the project — its aims and its methods — which support informed decisions at the end. The absence of such mechanisms was a clear factor in the relative lack of use of the Philippine evaluation in decision making.

Where donor and implementing agencies are involved, mechanisms need to be in place in both. They were lacking on both sides in the radio pilot effort.

5. Agency roles should be clearly defined.

In any situation with different agencies sharing responsibility, roles and relationships should be clearly defined.

It would be simplistic to say that a single implementing agency should always control evaluation, or that evaluation should always be separately managed. It does seem clear that these role relationships should be well-specified and understood.

This is particularly true where a donor agency with technical competence chooses to be involved in technical and substantive decisions. The influence of the funding agency should not be underestimated. If the agency chooses to have a substantive role it should accept the concommitant responsibility for continuity, leadership and, eventually, project outcomes. It should be especially careful in situations where there are multiple implementing agencies, reinforcing established role relationships.
6. **Implementation needs as much attention as evaluation.**

And perhaps more. It may be that summative evaluation should not be attempted until all elements of the pilot have been demonstrated to function effectively. In the Philippine case, for example, summative evaluation and the initiation of broadcasting commenced at the same time. Weak broadcasting led to a weak treatment, contributing to ambiguous outcomes. Problems in procurement of vehicles and other equipment in the Philippine project, if solved prior to pilot year implementation, would have had less serious impact on project implementation.

This approach would lengthen pilot projects somewhat, of course, unless clear focus and simplification provided counteracting time savings.

**SUMMARY FOR THE BANK**

This study was prompted by concern in the Bank that it might not be "... the ideal agency to carry out small scale, innovative projects."

The study suggests that this is true to the extent that Bank organization and procedures make it difficult to observe the lessons derived from the case. There are aspects of Bank operations which, on the surface at least, would seem to make pilot projects problematic. The size of such projects may be such that they must always receive secondary attention. The continuity and depth of involvement of Bank staff required may be difficult to achieve. The high level of technical sophistication of Bank staff may lead to overly complex designs and overly ambitious projects. It may also support a tendency to seek widely generalizable findings at the expense of focus on country or situation-specific learning.

On the other hand, there is no reason why the Bank could not deal successfully with such issues. The choice would seem to revolve not so much around capability to support pilot projects, but rather around the degree to which the Bank, as policy, seeks to take an active role in generating innovative solutions to development problems. If it chooses to be active, then pilot tests of innovations are an important tool to that end, and the Bank should organize accordingly.
NOTES

Introduction and
Chapter One: Project Initiation


3. Ibid., Appendix 2.


5. Ibid., p. 2-8.

6. Ibid., pp. 2-10, 11.

7. Ibid., pp. 2-12, 13.


Chapter Two: Formative Year


5. Ibid.

Chapter Three: Pilot Year


5. Interviews, Evaluation Director, Pilot Project Director, ECO Administrative Staff, January, 1984.


12. Interviews, EDPITAF Director, Pilot Project Director, Evaluation Director, January, 1984.


17. Memorandum, EDPITAF Director to Regional Education Directors, March 20, 1981.


Chapter Four: Endgame

1. Interview, EDPITAF Director, January, 1984.


6. ECO Staff memorandum, January 24, 1984.


10. Interview, EDPITAF Director, January, 1984.

NOTE ON CASE STUDY METHODOLOGY

John Middleton
School for International Training
Brattleboro, Vermont

December 31, 1984

Case study is a form of naturalistic inquiry useful in studying complex social processes from a wholistic perspective. The method involves the study of processes as they occur in the social environment, without intervention of the researcher either to control variables or to impose predetermined units of measurement of outputs. The purpose of the method is to increase understanding, not to verify prior hypotheses and propositions. As such, the method is well suited for a variety of purposes where planners and decision-makers are concerned with learning about process and context.

This summary description of case study methodology indicates this research approach is appropriate to the study of the Philippine Radio Education Pilot Project. A brief review of the methodology employed may help readers evaluate the study.

GENERAL METHODOLOGICAL APPROACH

Unlike positivistic research, which seeks to verify the "facts" of a social process as scientifically verifiable elements of reality, case study is concerned with understanding social reality from the points of view of people engaged in the process under study. Thus the methods employed are concerned with establishing rigorous procedures for identifying key actors, for gathering and interpreting data on the process in question and its context in place and time, and with verifying the explanations that emerge from the study with key actors. As with much social science research, the methodology of inquiry is reflected in a set of procedures.

Procedures

There are five phases to case study research, each involving certain procedural steps. These phases are briefly summarized below, together with comment on how the procedures have been applied in the present case study.
1. **Conceptualization**

An early step in developing a case study is conceptualization. The researcher defines the boundaries of the system to be studied. This can be challenging. Unlike the survey researcher, whose data gathering can be limited by theoretical constructs of variables, the case study researcher cannot limit the research to a pre-determined set of specific variables. Instead, he/she is confronted with the possibility of dealing with an ever-increasing set of processes and variables as more is learned about the phenomenon under study and its relationships with its environment.

In the present study, the boundaries were relatively easily defined as encompassing the actions and interactions of the Bank and relevant Philippine agencies in connection with the radio education pilot project. These boundaries not only limited the study in terms of the organizational set, but also set time boundaries. Interestingly, it was found helpful to go back in time some three years prior to the official beginning of the project in order to adequately define the organizational and political context of the pilot, and to review the post-pilot period to fully trace and understand organizational actions during the pilot itself.

Case study researchers like to debate the point at which specific variables begin to be defined. Some methodologists hold that studies should begin without pre-determined concepts, with variables emerging from the data. Others hold that, certainly in evaluative cases, pre-existing questions and concerns will exist and can guide the study.

In the present case pre-existing Bank concerns over the apparent lack of impact of the pilot project in decision-making guided both the establishment of the boundaries of the study and the investigator's work.

2. **Data Generation**

Case studies can draw on any and all data. Primary sources are the key actors in the process being studied. Direct observation of key elements in a process can be important, as can analysis of documents and existing statistical data. The Philippine case draws on each of these kinds of data.

Interviewing, of course, is a key technique in developing a data base. Given that case research aims at expanding understanding of a complex process, interviewing strategies in early stages of a study are generally quite open ended, with the intent of inducing key actors to tell their story with as little promoting as possible. As the study proceeds, and as an increasingly complete picture of the process develops, interviewing becomes more specific as particular facts, events or interactions become significant in the analysis.

3. **Constant Comparison, Inductive Analysis and Triangulation**

The researcher interacts with the data as it is gathered, formulating and reformulating categories of knowledge and relationships. These categories and patterns emerge as understanding grows, and must be
constantly reformulated to fit with the emerging case. This process of constant comparison and inductive analysis is the methodological counterpart of the notion of beginning the study with relatively few pre-defined concepts.

In the Philippine case, initial data gathering and analysis was relatively unstructured. The Bank concern for the use of the pilot project in decision-making was the principal guiding concept.

The internal reliability of the case is primarily determined through triangulation of information. To the extent that a pattern is supported by data from several sources it may be considered relatively reliable. Divergence of data on a given point is, of course, a clear signal to the researcher for more investigation.

It is important that the researcher constantly check ideas, insights and conclusions being generated with key actors in the process. The goal is to get agreement among key actors regarding the factual elements of the study. Ideally, this is a constant process as tentative conclusions emerge from the study. In concrete terms, it requires sharing of interim research notes, preliminary concepts, and, eventually, successive drafts of the study. Where full agreement cannot be reached on interpretation of events (as is often the case), the researcher assumes final responsibility.

The geographic distance between key actors in the Philippine case (Washington, D.C. and Manila) has limited the extent of interaction possible. There were two rounds of interviews in the Bank, and one review of a first draft of the study. A second draft was reviewed by colleagues in the Philippines and by consultants to the Bank who played active roles in the radio pilot project.

4. Building a Data Base

Constant inductive analysis requires a growing base of organized information for the analysis. The researcher must give considerable attention to the data base of the study.

In the present case, permanent records of interview notes (usually dictated within two hours of interview completion) have been kept. Various Bank reports have been formally analyzed. The flow and timing of key events has been determined through retrospective scheduling techniques.

5. Preparation and Negotiation of the Report

The final version of the study, in keeping with the process of inductive analysis, has been "negotiated" with key actors in order to reach as much consensus on fact and interpretation as possible. This can require an extensive process of review and revision of successive drafts of the study, but is important to internal and, eventually, external validity.
GENERALIZATION FROM A CASE

In much social science research, generalizability is determined by the adequacy of the methodology employed. There is tacit agreement between the researcher and the users of research on the degree to which a particular research design and statistical analysis can be trusted to yield objective truth. With case study, generalizability rests to some extent on the methodology — particularly triangulation and the support of conclusions with a variety of evidence. More importantly, generalization rests on the appeal of the data to the reader's own knowledge and understanding.

There are generalizations of two kinds. The first, generalizations about the case, lead the researcher to specify patterns and conclusions from the data and to support both with evidence. Chapter Five of the present study exemplifies generalizations of this type.

The second is generalization from the case to other situations. Use of the concepts and patterns from one case to illuminate or guide action in another is, in the final analysis, the responsibility of the reader. It is for this reason that case studies should be rich in the detail which helps the reader compare the context of the case with the context of the new situation as a central process in determining the relevance of the case for the situation at hand.
SECTION II:

EVALUATION OF THE PAKISTAN PRIMARY EDUCATION PROJECT:

A METHODOLOGICAL CASE STUDY

Paul A. Schwarz, Consultant
Education and Training Department
World Bank
Preface

Over the past decade, as experience has been gained with the use of evaluation as an aid for decision-making, the effectiveness of traditional evaluation paradigms has been called into question. This paper, the first in a series on evaluation components in education projects financed by the World Bank, examines the question of appropriate evaluation design, using the Pakistan Primary Education Project as a case. The paper distinguishes between educational activities that are fully development and understood and pilot activities, and explores the implications of this distinction for choosing evaluation questions and methods.

The objective of the Primary Education Project in Pakistan, appraised by the Bank in 1978, was to pull together the most promising of the educational reforms that were then being championed, introduce them on a trial basis, and put them to the test. On the basis of the results, the most effective of the reforms would be implemented on a larger scale in a second, follow-on project. The major product of the initial project was to be a "comprehensive proposal for the expansion and improvement of primary education" based on empirical data. Thus, the role of the evaluation in this case was to help predict the effects of future versions of the reforms that were being given a trial under the present project.

During supervision of the project, the author examined the evaluation activities that were designed for the project, and concluded that many assumed that a stable intervention was in place, rather than a set of activities that were changing constantly as more was learned about what worked and what didn’t. The paper describes the actions the author took midway through the project to reorient the evaluation activities towards the decisions that needed to be made, actions he calls collectively "reconnaissance." He describes a variety of methods that were used to collect systematic information during the course of interviews and site visits.

In a final section, the author assesses the applicability of the work in Pakistan to other projects. He recommends flexible project design and management that would allow changes in project activities as information is collected about how things are working. Among the techniques the recommends for providing timely and useful information are (i) the use of diaries and other record-keeping procedures at field sites; (ii) periodic reconnaissance visits; (iii) formal annual review meetings of all concerned parties; and (iv) special studies on topics identified by the foregoing procedures.

Paul A. Schwarz
World Bank
September, 1983
Evaluation of the Pakistan Primary Education Project:

A Methodological Case Study

The most highly developed techniques of program evaluation are those based on the design of experiments. They are rigorous and reliable, and can measure even small effects. But they have been greatly overused. In the field of education especially, they have been applied to situations for which they were not intended, and in which the results they yield can be meaningless. This case study illustrates an approach that provides more useful feedback on the merit of innovative programs.

The account begins with a brief overview of the methodological issues. It suggests that development projects can be divided into four different groups from the standpoint of their evaluation requirements. The "standard" evaluation design fits only one of these groups. For projects in the other three, specifically tailored designs have to be developed. It describes the process that was followed in Pakistan to develop an evaluation design for a project that is representative of the largest group. It concludes with a number of procedural suggestions for other projects of this type.

The Methodological Context

The techniques of experimental design were developed for the evaluation of innovative agricultural treatments—a new nutrient formula, a more intensive spraying schedule, a change in the spacing of plants, etc. The objective was to establish whether the experimental treatment led to better yields than the existing practice. If so, it would be adopted. "What yield can be expected if the identical treatment is administered again under the identical conditions?" Was the question that the experimental techniques were designed to answer. Assessments based on these techniques predict the effects of identical replications.

The most extensive of the nonagricultural applications that have been made of experimental designs has been in the testing of pharmaceuticals. Here too the "independent variable" is a treatment that is to be administered in accordance with rigid specifications, and that is to take the identical form each time it is applied. Here too the assessment is limited to the treatment in its present form. An experimental drug that is only one modification removed from a medical breakthrough is rejected as forcefully as a drug that is totally wrong in principle. Experimental designs are limited to go or no-go decisions about the treatment as it is at present.

The nearest analog in the field of education is the evaluation of books, films, computer-based instructional programs, and other "treatments" that are invariant from use to use. To predict the sorts of results that will be obtained from the introduction of a new text or title, a controlled experiment is by far the most powerful of the available techniques.
Exercising adequate controls (e.g., with respect to such potentially confounding variables as the quality of the teachers) may be more difficult in educational settings than in agricultural or pharmaceutical research. But the experimental methods clearly are the methods of choice for assessing invariant inputs or processes when the logistics can be managed.

We shall characterize these kinds of go or no-go assessments as tests of readiness. They provide a forecast of the benefits a trial procedure can be expected to yield if it is replicated precisely.

In most educational development projects, the readiness of the activity for operational use is not the central issue. Usually, the assessment is made while the activity still is in a formative stage. The issue is its potential. Do the early indications suggest that this is a promising approach? Are further investments in refining and improving it likely to pay off? There is no thought of adopting the treatment in its present form, and certainly not of repeating the mishaps and miscalculations of the trial phase. A forecast of the benefits that can be expected from a faithful replication would be meaningless. The task is to predict what a refined, still-to-be-developed version might be able to accomplish.

The inappropriateness of experimental designs for the evaluation of programs still taking shape has not been widely recognized. The conventional wisdom has been to use experimental or pseudo-experimental techniques whenever practicable. The typical result has been overwhelmingly negative findings. In the evaluation of the Great Society programs in the U.S., this was one of the consistent oversights that made all but inescapable the uniform findings of failure.

A second, related oversight has been not to differentiate adequately between improvements that are necessary to achieve a certain result and improvements that in and of themselves are sufficient to produce it. The criteria used to evaluate a program typically have been based on its ultimate benefits; e.g., on the gains in the achievement of the students exposed to it in the case of education. This is appropriate for programs that are intended to provide the students with a complete learning environment, such as the IMPACT system developed in Southeast Asia. But programs so comprehensive in scope are the exceptions. More commonly, interventions are limited to a certain aspect of the learning environment. They are intended to be a part of the solution, not a total solution. Evaluating them as though they were total solutions obscures their true effects.

Evaluations of in-service teacher training programs are a case in point. In most developing countries, and especially in rural areas, the need for more highly skilled teachers is not debatable. It is difficult to imagine a solution that does not include improvements in the teacher service as a central component. Yet, assuming that the effectiveness of an in-service training program will be revealed by gains in student performance is risky. Any number of factors — an incompatible curriculum, inadequate materials or facilities, resistance to change on the part of the principals — can preclude impact on student achievement. The effects of a
good program and a poor program can appear to be the same. The appropriate measure of the effectiveness of a teacher training program is the improvement that it produces in the skills of the teachers.

It follows that the design of an evaluation should begin with two fundamental questions about the status and scope of the project:

1. Is there reason to believe that the treatment is ready to be replicated in its current form, or is the issue its potential?

2. Are the improvements that the project is designed to bring about sufficient or only necessary conditions for producing the benefits sought?

There are four different situations, as summarized in Figure 1. Each requires a different approach.

For projects that fit situations of Type IV, the approaches emphasized in the texts on program evaluation are appropriate and powerful, as noted earlier. Because the issue is readiness, experimental designs will provide the information sought. Because the inputs are intended to be sufficient to produce the desired effects, using measures of ultimate effects as the criteria is fair. As a practical matter, it often is desirable to include a number of additional, diagnostic measures, however, to help formulate next steps if the solution being evaluated turns out to be ineffectual.

Projects that fit situations of Type III also are amenable to experimental approaches provided that the criterion measures are adjusted. Because the program does not try to meet all of the prerequisites to the achievement of the desired outcomes, the use of ultimate criterion measures would be misleading. This problem was discussed above in the context of in-service teacher training. The appropriate criteria are measures of the "intermediate" outcomes (such as improvements in the teachers' performance) that the project can be expected to achieve within the scope of what it does. The results of this sort of intermediate evaluation indicate whether or not this treatment is ready to be adopted to meet this aspect of the need.

Investments in Type III situations cannot stop with the decision to adopt the trial procedure, of course, even if the evaluation shows it to be effective. The goal is to produce the ultimate outcomes. The remaining needs also must be identified and met. For this reason, it generally is desirable to include additional, diagnostic measures also in Type III situations. Their role is to help answer the post-evaluation question of: What else must be done to move beyond these gains to the ultimate objectives?

<table>
<thead>
<tr>
<th>Test of Potential</th>
<th>Test of Readiness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Necessary Inputs</td>
<td>Type I</td>
</tr>
<tr>
<td>Sufficient Inputs</td>
<td>Type II</td>
</tr>
</tbody>
</table>

Type III

Figure 1. Different evaluation contexts
For projects that fit situations of Types I and II, there is no established, coherent approach. Nor can there be a fixed all-purpose model. The task is to assemble evidence about the project that will enable the responsible officials to determine whether or not to invest in its further development; and, if the decision is affirmative, what the next steps should be. The nature of "best evidence" varies with the nature of the inputs, the anticipated outcomes, the alternatives that are available, and the political surround. It differs from project to project. Textbook designs may provide useful ideas, but they cannot be applied directly.

The Primary Education Project in Pakistan consists of a number of interventions of Type I. Its objective is to provide a basis for assessing their respective merits, so that the most promising can be combined into a comprehensive solution of Type II. We describe the project below.

The Primary Education Project

The project appraisal was carried out during August-September 1978; the Staff Appraisal Report is dated March 15, 1979. It identified four basic problems as follows:

1. Unequal Opportunity. Though primary school enrollment in Pakistan had been rising steadily (from 40% to 54% since 1966), there were large inequities in access to schooling both by region and gender. Enrollment rates across provinces varied from 30% to 59%. The rate for boys was 73%, as contrasted with only 33% for girls.

2. Wastage. Dropout rates were high. About half of the students who entered Grade I dropped out before Grade 5. Here too there were large imbalances by region and by gender. In the least advantaged province, the drop-out rate was 64%. About 40% of the girls who entered Grade I dropped out before Grade 2.

3. Low Quality. A major reason for the high rates of wastage was the poor quality of primary education. Physical facilities were inadequate; instructional materials were few and inappropriate; large numbers of teachers had no professional training; teacher absenteeism was rampant.
4. Financial Constraints. Government spending for primary education in Pakistan remained below the median for LDCs even though the allocation had been increased in the most recent development plan. There were not sufficient resources to meet existing needs. Significant increases in enrollments and retention might overwhelm the budget. More economical ways of delivering primary education had to be introduced alongside the qualitative improvements. These four problems were the problems that the project was intended to address.

An intervention of approximately five and a half years duration was planned. The tentative schedule was as follows:

- Jul 1979 - Feb 1981: Preparations
- Mar 1981 - Feb 1982: First experimental year
- Mar 1982 - Feb 1983: Second experimental year
- Mar 1983 - Feb 1984: Third experimental year
- Mar 1984 - Sep 1984: Analysis of results

The goal was to evaluate a sufficient number and variety of inputs to design a comprehensive program for the improvement of primary education throughout the country.

The Appraisal Report identified four broad categories of inputs that would be included in the project. Each was expected to exert leverage on one or more of the target problems.

The first was the construction of physical facilities, and the provision of related equipment and furniture. Three different kinds of facilities were included in the initial plans:

1. A total of 1,014 new classrooms would be constructed, equipped, and furnished. Multiple benefits were expected. By providing additional facilities, new classrooms (a) would help to mitigate the problem of access. Because they would be more attractive than existing classrooms, they also would attack the problems of (b) pupil drop-out and (c) teacher absenteeism, and (d) be conducive to more effective instruction. Because most of them were to be additions to existing schools rather than free-standing, they further (e) would permit economies of scale that would reduce per capita costs.
2. A total of 5 in-service training centers would be constructed, equipped, and furnished. These would support the in-service teacher training component of the project (as described further below).

3. A total of 288 residences for female teachers would be constructed and furnished. They would be located in rural areas where the lack of qualified female teachers is a major barrier to the enrollment of girls. The expectation was that adequate residential facilities would make it possible to recruit qualified female teachers from urban areas.

At a later stage in the planning of the project, provisions for the construction of 182 boundary walls for girls' schools were added. This was because the absence of boundary walls (and the resulting lack of privacy) can be a major factor in the parents' decision not to send their daughters to school. The number of residences for female teachers also was increased (to 334) after the staff appraisal.

The second was the improvement and expansion of the supervisory system. Supervision in Pakistan is the responsibility of Assistant Education Officers (AEOs) who are expected to make regular inspection visits to all of the country's schools. But, because the typical AEO is expected to cover 150 or more schools, generally without transport, visits tend to be rare. The primary schools of Pakistan are largely unsupervised. The project undertook to correct this in a sample of 26 districts.

The plans called for the introduction of one or two new supervisory positions in each of the participating districts. The new position of Learning Coordinator would be established in all of them. A total of 470 Learning Coordinators would be deployed, each to take responsibility for regular visits to a specified caseload of schools. Two of the provinces (encompassing 18 of the participating districts) would establish also the additional new position of Supervisor. A total of 50 Supervisors would be deployed to exercise oversight over the Learning Coordinators in these two provinces and generally manage the program.

The plans also called for deliberate variation in the number of schools for which Learning Coordinators and Supervisors would be assigned responsibility. This would permit comparative assessments to establish the most cost-effective caseloads. Multiple benefits were expected. By supervising and helping the teachers, the Learning Coordinators and Supervisors would upgrade the quality of instruction. By working with parents and community leaders, they would promote enrollment and cut wastage. To insure that schools visits could be made frequently enough to produce these benefits, explicit provisions were made for transportation. Male Learning Coordinators would be provided with motorcycles; female Learning Coordinators would be provided with vans.
The third was in-service training of current teachers, and recruitment of additional teachers, to strengthen the teacher service. In-service training courses 2-3 weeks in length would be provided to approximately 10,000 participating teachers before the start of the program. Additional recurrent training courses would be provided throughout the three experimental years. More than 200 teacher trainers would be selected and trained to implement this component.

The additional teachers that would be recruited would not be regular (i.e., certified) teachers. Instead, the new position of Assistant Teacher would be created. Assistant Teachers would have at least a middle school education, and would teach in schools near their homes. A total of 340 (more than half of them women) would be appointed. The expectation was that "the use of these less qualified teaching staff...would be)...promising cost-reducing factor and a means of supplying female teachers in rural areas" (SAR, p.20). Though they lack formal teaching credentials, "the performance of such personnel is often superior to that of "regular" teachers, because assistants are employed near home, have better rapport with pupils and parents and understanding of local needs and tradition, and are more appreciative of the chance for cash employment" (p.18).

The fourth was the provision of instructional materials. Each of the participating schools would receive one set of library books, one set of teachers' guides, and one set of general, agricultural, sports, and social studies equipment. Each participating teacher would receive one set of textbooks. Ten percent of the participating pupils in one province would receive free textbooks in the first year of the project; all of the participating pupils in a second province would receive free textbooks during all three years of the project.

A number of other inputs was designated as appropriate for inclusion in the government's overall strategy for improving primary education, but omitted from the project to keep it to manageable proportions. These included salary adjustments for teachers, free school meals, bilingual curricula, and self-help construction.

Overall, the project "would include more than 4,000 schools, nearly 10,000 teachers, and over 380,000 presently enrolled students and about 400,000 children who are not now enrolled...[It] would be limited to rural areas; it would cover 7-8 percent of the nation's primary schools, teachers and enrolled children" (p.18). The total cost was projected at US$17.2 million. As basically a design and planning study, the Primary Education Project in Pakistan represents a large, perhaps unprecedented, investment in getting pertinent facts.

As an important by-product of this investment, the Staff appraisal also envisioned the development of greatly strengthened Pakistani capabilities in project management and (especially) in evaluation and research. To this end, a three-month training program for a cadre of administrators and researchers was added to the other training provisions. A sizable technical assistance component, consisting of both local and expatriate consultant services, also was included.
The Research And Evaluation Component

The answers to the questions that were raised in the introductory remarks -- i.e., potential vs. readiness and necessity vs. sufficiency -- were explicit in the Appraisal Report. Despite the magnitude of the project, it was not viewed as a pilot study of a solution that would be replicated if it proved effective. The pieces were not expected to add up to a coherent whole. The "central purpose... (was) ...experimentation into the contribution of selected inputs to achievement of project objectives" (p.26). On the basis of the findings, some subset of the inputs would be combined into clusters that might be sufficient solutions. But this would be done in a follow-on phase, not as part of the initial project.

In the terminology introduced earlier, then, the project was to be a type I enterprise, the chief purpose of which would be to permit separate (Type I) assessments of a number of different inputs. On the basis of the findings, a follow-on project would be designed to implement a more comprehensive (Type II) solution. The major written product was to be a "comprehensive proposal for the expansion and improvement of primary education based on the...results of the project" (p.27).

This clear view of the project and its goals was not mirrored in the proposed evaluation methodology, however. In 1978, the commitment to the experimental techniques was at its peak. The accepted approach was to try to fit all projects into this model to the extent this could be done. The approach proposed in the Appraisal Report was based on the collection of data on three kinds of ultimate outcomes,

1. rates of enrollment and drop-out,
2. level of pupil achievement, and
3. cost;

and on three kinds of intermediate outcomes,

4. pupil and teacher attendance,
5. teacher knowledge and competence, and
6. attitudes of pupils, parents, and community members.

Baseline data were to be collected before the start of the first experimental year. A group of control schools was to be included.

These plans were changed substantially after the staff appraisal. Measures of costs, of pupil and teacher attendance, and of teacher knowledge and competence were dropped from the plan. The attitude measures were expanded to include also the attitudes of teachers. Descriptive measures of a wide variety of school and community characteristics were added. The revised plan was to carry out annual surveys of

1. pupil achievement,
2. pupil, teacher, and parent attitudes, and
(3) selected demographic characteristics, including enrollment and wastage.

The data would be collected by the Learning Coordinators in the course of their visits to schools. No control groups would be used.

A contract for technical assistance was executed with the Scottish Council for Research in Education (SCRE) through the British Council. Consultants from SCRE worked with Pakistani specialists on the development of the data collection instruments necessary to implement these plans. In addition, they recommended the inclusion of

(4) "illuminative" studies of important process elements,

and this recommendation was adopted. Plans were made for illuminative studies of Learning Coordinators, Assistant Teachers, dropouts, textbooks, buildings and furniture, parental perceptions, and teacher training.

The management of the research (and, indeed, of the project as a whole) was largely decentralized to the four provinces. Though responsibility was vested in a Federal Implementation Unit (FIU), this was primarily a coordinating mechanism. Within the broad outlines of the overall project plan, the Provincial Implementation Units (PIUs) exercised full decision-making authority. Three of the four PIUs in turn delegated substantial responsibility for the research and evaluation component to Institutes of Educational Research and other institutions. The fourth maintained full responsibility for carrying out this component in-house.

The collection of baseline data on the achievement, attitude, and demographic measures was completed in May 1981, a few months after the beginning of the first experimental year. Because of the variety of administrative and technical problems, however, the project staff had substantial doubts about the accuracy of these data, and decided not to use them. Instead, it was decided that the initial data collection would be treated as a practice run. The second year's survey would by used as "baseline" data to establish the magnitude of the changes the program produced.

The Evaluation Component Rethought

The first outside review of these plans was carried out as part of a Supervision Mission in September 1981. The following problems were noted:

1. The data being collected on ultimate outcomes — enrollment, wastage, and pupil achievement — were unlikely to show palpable changes during the course of the project. The inputs had not been clustered in ways that might make them sufficient to produce such changes. This was not the intent in Phase 1. Should changes occur, there would be no way to attribute them to the intervention, moreover. Indeed, the more
parsimonious assumption would be that they were the result of other factors. Neither the achievement testing nor the demographic survey could be expected to contribute to the comprehensive proposal.

2. The attitude tests might have been designed to make certain, albeit small, contributions to the proposal. In fact, there was little apparent relationship between the items on these tests and the goals of the project. The teachers' agreement or disagreement with such statements as

- Teaching is never a tiresome job,
- Children should be allowed to choose the seat of their choice, or
- An aggressive child is a great problem

would not speak to the utility or in utility of any of the interventions. The results might be interesting intellectually, but, like the data on achievement and demography, they could not be expected to contribute to the comprehensive proposal.

3. The relevance of the illuminative studies could not be evaluated because work on these had not begun. But there was no reason to suppose that they would have the proposal-specific focus that the other measures lacked. As their name implies, the purpose of illuminative studies is to contribute to knowledge. Their goal is to transcend, not to support, the practical goals of an action project. The more or less haphazard selection of topics on the basis of their intrinsic interest rather than their responsiveness to the needs of the project had been consistent with this orientation. Unless steps were taken to reorient these studies, they too would fail to contribute to the comprehensive proposal.

Overall, the odds-on prospect was that the project staff who would be assigned the task of drafting the proposal would find that the reams of print-outs produced by the evaluation component contained no facts that they could use. A five-year investment in generating facts for the design of a large follow-on initiative would yield nothing of practical value. The mission recommended two immediate procedural changes.

The first was to restructure the evaluation component to give priority to the practical needs of preparing an action proposal. The Mission suggested that a rough outline of the important topics to be covered in the proposal should be prepared; that the information needs associated with these topics should be identified; and that the plans for data collection should be revised to fit this specific set of needs. Each of the data points in the revised design should be linked clearly and importantly to one or more of the topics to be addressed in the proposal. The studies had to be "prescriptive" rather than merely descriptive or informative to meet the needs of the project.

The second was to centralize the management of the evaluation component and to upgrade the research staff. Generating prescriptive data
required a unified design that would permit aggregation and comparison across provinces. The existing, decentralized approach would not result in data that could be combined. Focused, disciplined data collection also required strong central leadership. The Mission suggested that one or more senior researchers (Pakistani and/or expatriate) should be assigned to the FIU, and that more of the positions of "research officer" in the PIUs should be filled with people with actual research skills.

These recommendations were accepted in principle, and by the time of the second external review (August 1982) considerable progress had been made. The staff of the FIU had been expanded to include a senior educational planner and an experienced methodologist, and highly qualified individuals had been recruited to fill both of these positions. A series of workshops had been held to identify practical data needs. Within each PIU, a "research cell" was being created. Mechanisms for carrying out cross-provincial studies also were in place. The stage was set for designing and implementing the needed prescriptive studies. But a year had passed. The first draft of the comprehensive proposal was to be completed within the next ten months. The Project Director requested the Mission's help in completing the revised design at the earlier possible time.

The Mission suggested two specific steps. The first was to replace the experimental approach that motivated the original plan with a developmental approach based on the following logic:

1. The design of the project was grounded in many assumptions about the problems, their primary causes, the nature of appropriate remedies, etc. These assumptions reflected the latest and best of the available knowledge in 1978. Had a comprehensive proposal been drafted then, it would have had to rely on the identical assumptions. They represented the state of the art.

2. As a direct result of the project, however, these assumptions now could be refined. In 1978, the effects of increased supervision could be no more than speculative, for instance. In 1983, the effects could be observed. The comprehensive proposal could be written with greater assurance.

3. The most useful function that the evaluation component could serve, therefore, was to compile the new information that the project had made available about important assumptions. At the moment, this information was implicit in the observations and
It was by providing data on phenomena on which adequate data were not available before the intervention that the Primary Education Project (and Type I projects generally) could contribute most directly and importantly to the development of effective solutions.

Because of the wide range of inputs included in the project, and the (both planned and accidental) variations in their delivery, the number of phenomena on which data reasonably could be collected was large. Doing a thorough study on each aspect of the project was not possible within the time and resources available. Choices had to be made. The Mission's second recommendation was to base these choices on data. Specifically, the Mission suggested a "reconnaissance" of field activities to try to identify the sorts of inquiries that it would be most productive to pursue over the next ten months to generate data for the proposal. Both recommendations were accepted.

Conduct of the Reconnaissance

The reconnaissance was carried out during the period August 2-18, 1982 by the senior educational planner who had been recruited to join the staff of the FIU (Professor Habib-ur-Rehman) and the Bank Consultant. Visits were made to project sites in all four provinces. A total of 69 "grass roots" participants in the project (Learning Coordinators, Supervisors, Assistant Education Officers, certified teachers and Assistant Teachers) was interviewed, in addition to project staff and researchers.

A structured interview protocol was developed at the beginning of the activity. It focused on the following topics:

1. Problems in providing quality education,
2. Characteristics of the inputs being made,
3. Apparent effects and accomplishments,
4. Important process elements, and
5. Important contextual factors.

As the interviews proceeded, and the nature of the most productive lines of inquiry became progressively clearer, questions were added and dropped. The following, selective summary describes the kinds of information that were developed about each of these five topics:

Problems in Providing Quality Education. The questions on problems were formulated with two objectives in mind. The first was to take advantage of the fact that the interviewees were not only participants in this one project, but day-in, day-out observers of all that transpired
in Pakistan's primary schools. Rather than encouraging them merely to echo the conventional wisdom about the major problems of the educational system (by asking them to name problems), the interviews focused on observations and experiences that would reveal the problems they actually encountered. The second was to try to go beyond the problems specifically addressed by the project to any others that the effort to solve these may have brought to light. Limiting the inquiry to the problems that the project was intended to address (as is generally done in experimental designs) would have run counter to the goal of refining the assumptions made in 1978 about the nature of the needs.

We used two techniques. The first was a "nominations" technique, in which we asked each Learning Coordinator and Supervisor to identify by name the best and worst (and next-to-best and next-to-worst) schools in their caseloads, and then asked questions about this specific sample of schools. "What makes this school worse than all of the other?" was a typical question. "In what other ways are these two schools (i.e., the best and the worst) different?" was a typical probe.

This approach quickly identified the single-teacher, multi-grade school as one of the principal problems. Again and again, the fact that there was only one teacher was cited as the main reason that the worst school in each set was so inadequate. Our respondents agreed that teachers with the minimal qualifications of teachers in rural Pakistan simply cannot manage five grades concurrently. When the teacher is ill or has to attend to personal needs, moreover, a single-teacher school down. This problem had been mentioned but not emphasized in the Appraisal Report. It had not shaped the project plans, nor had it been included in the topics to be explored with illuminative studies. It had not been given nearly the attention that our respondents' observations suggested in deserved.

As this began to be apparent, we began to collect data on the numbers of teachers and pupils at each school. These confirmed the extent of the problem. We found that the proportion of single-teacher schools ran as high as 65 percent of boys' schools and 85 percent of girls' schools in the least advantaged province, and from 30 to 50 percent in the other three provinces. That the interventions being tried were in and of themselves adequate to make significant inroads into this problem seemed doubtful. More would have to be done.

(1) The approach that will be taken in the follow-on project to improve the quality of education in single-teacher schools will have to be one of the major topics discussed in the comprehensive proposal. The collection of data that could help to formulate a persuasive approach should be given high priority in planning the activities of the next ten months.

Our data also revealed that most single-teacher schools are small. Though they generally had students in all five grades, there often were no more than one or two students in Grades 4 and 5. Total enrollment
might be fewer than 20. This complicated the task. Upgrading so many small schools, if it could be done at all, would be enormously expensive. To be responsive, the proposal would have to discuss the possibilities for school consolidation; and, even though this is a policy decision, not a topic for research, it would be important to document the approach proposed with supportive data.

(2) Data on the feasibility of alternative approaches to the consolidation of small rural schools would provide another, extremely useful resource for preparing a credible proposal.

A further complication that was inherent in these findings was that such high proportions of small single-teacher schools cast doubt on the adequacy of the pupil/teacher ratios that were being used as targets for teacher recruitment and training. Bank projects in Pakistan use of target of 33:1, for instance. Is this a useful guide? In a single-teacher school with 22 pupils, it implies a surplus of 50 percent in the teaching staff. A large number of schools with "surpluses" of this magnitude clearly will lead to large underestimates of the actual needs for teachers.

(3) Proposals for increasing the teacher service should not be based on gross pupil/teacher ratios. At a minimum, single-teacher and multi-teacher schools should be analyzed separately to establish and forecast needs.

When we applied this rule to our own data base, we found that the multi-teacher schools in three of the four provinces were reasonably close to the 33:1 target. Except for some distributional imbalances, these provinces had an adequate supply of teachers. In the fourth province, this was not the case. Even with maximally efficient redistribution, the teacher service in this province would have to be increased by 35 percent to achieve the target ratio. In the short term, this could not be done.

(4) In addition to the steps proposed to address the problem of single-teacher schools, the comprehensive proposal also will have to address the large, overall shortage of teachers in this province.

The second technique we used to elucidate problems was to obtain a "time sample" of the activities of the Learning Coordinators. We asked each of them to tell us which school or schools they visited on certain (arbitrary) dates, and what they did during the course of these visits. Establishing the kinds of help that they were providing most frequently would indicate the kinds of needs that they were encountering most frequently, and thereby the most prevalent problems. Not all of the Learning Coordinators maintained written logs of their daily activities. But the reports of those who did and those who did not were reasonably consistent. Two further problems emerged.
One was that more of the teachers they worked with did not have sufficient mastery of the subject matter, even at the level of the primary school curriculum. Especially in math and science, they did not understand concepts and techniques they were supposed to be teaching their students. The Learning Coordinators regularly spent time helping the teachers with the more difficult topics. In the in-service training sessions we observed, the difficulties the teachers were having with substance also were apparent.

The implications were similar to those of the problem of school size. Accepting the content of the current curriculum and trying to raise the skills of the teachers to the requisite level would not be cost-effective. The possibilities for simplifying the curriculum had to be considered first. At least superficially, the prospects for trimming seemed good. A number of the leading educators in Pakistan have long held that the emphasis on the "new math" and the more theoretical aspects of science are inappropriate in rural schools, and these are the same elements that the teachers are least well prepared to teach.

(5) Alongside its data collection activities, the project staff should lobby for timely decisions on curriculum reform over the next ten months, to try to reduce the shortfall in teacher skills as much as practicable before having to take a position on investments in teacher training in the comprehensive proposal.

The other problem that emerged from these data was that the predominant style of teaching remained the teacher monologue. Even such simple suggestions as using pebbles or leaves to illustrate the addition of sets were viewed by some teachers as heresy, according to our respondents. The in-service courses had emphasized the use of home-made teaching aids. The Learning Coordinators were trying to promote this idea as part of their school visits.

This problem differed from the others in that it seemed to provide an opportunity for fairly substantial near-term improvements. Unlike the management of a multi-grade school or the teaching of math and science, the effective use of teaching aids can be taught fairly easily and inexpensively.

(6) Because of the substantial gains that can be promised with fair confidence, the steps that will be taken to improve pedagogical style should be developed in detail in the comprehensive proposal.

The data we collected suggested a number of specific possibilities in this regard, as elaborated later in the discussion.

Characteristics of the Inputs Being Made. In experimental designs, the collection of data on inputs usually is limited to the
information needed for management purposes. Only the bare facts about quantities, costs, and delivery schedules are recorded. Because the use of experimental designs presumes a fixed treatment that is faithfully replicated at every site, cataloging the inputs that are made is not expected to be informative. Quality control mechanisms to guard against slip-ups suffice.

In Type I projects, variations in inputs must be anticipated. Establishing exactly what was (or was not) done at every site is fundamentally important. Data on the outcome at a given site cannot be adequately interpreted without knowledge of the inputs. Variations in inputs also constitute prime learning opportunities for improving the project in that they reveal the differential effects of different delivery methods.

We used no special techniques to establish the status of the inputs. We merely probed for details. We tried to pin down exact numbers, locations, and dates. We asked for exact descriptions. When possible, we observed the process ourselves. We found large and significant deviations from the stated plans that had to be taken into account in the design of evaluative studies.

In the province in which we began, for instance, we found that few if any of the 300 new classrooms being built would serve the purpose envisioned in the Appraisal Report of increasing school size to reduce per capita costs. More than half of them (180) had been set aside for use by the Learning Coordinators as "learning resource centers" for meetings of teachers, etc. Most of the remainder had been built as free-standing one-room schools in locations that had no physical facilities whatever, and held classes out of doors. In principle, this policy of meeting the needs of the neediest made good sense. But it was not consistent with the project's goals. It tended to perpetuate rather than resolve the problem of small schools.

(7) Interpreting the effects of the construction component in this province as the effects of building larger schools would be misleading. This is not how the money was spent.

Nor was this deviation from the original concept limited to this one province. Similar decisions were made in the other provinces. The design proposed in the staff appraisal had missed two important points. One was that the existing facilities are so grossly inadequate for even existing enrollments that expanding them with a view to increasing enrollment is not realistic. The data we collected on nearly 500 schools suggest that 40-50 percent of the schools in rural areas have either no facility at all or just a single room. Government cannot afford to catch up with the existing demand, much less consider expansion. The other is that many of the "classrooms" that seem to be in place are not. What appear to be classrooms on paper in fact may be loaned or rented rooms, or a corner of the community meeting hall, or even the structure that houses the village pump.
Pakistan appears to be a land of far too many schools and far too few classrooms, and this reality will have to be reflected in the discussion of the construction component. The original plan will have to be revised.

From the standpoint of the proposal, the enormity of this need seemed to represent more of an opportunity than a difficulty, however. If Pakistan sooner or later will have to make a large investment in school construction, establishing the sorts and sizes of facilities that are the most efficient clearly is important. Because of the variations in the implementation of the construction component in the four provinces, this project (and perhaps only this project) can provide directly relevant data. One province, for instance, had built not only one-room schools, but also a fairly large number of two-room schools; and, at some locations, had expanded existing two-room facilities to four. There were ready-made opportunities for assessing the advantages and drawbacks of alternative designs.

The collection of the data on the effects of school size that are potentially available at the project sites may point the way to substantial economies in the huge investment in school construction that Pakistan is facing, not only in the follow-on project, but for the foreseeable future.

Sizable deviations from plans occurred also in the supervision component, but for different reasons. Here, the provinces did not decide to change the initial plans deliberately, as they had in the case of the construction component. The deviations that occurred were not intended. But in one province in particular, they nevertheless were large. Because of a series of unanticipated events, the amount of additional supervision that in fact was provided in this province in the 18 months since the project began was essentially zero.

Five separate factors had combined to undercut the supervision arrangements in this province. One was that the intended limits on caseloads had not been observed. Instead of the maximum of 21 schools that was to be assigned to male Learning Coordinators, the actual maximum crept up to 37. Instead of the maximum of 17 schools to be assigned to female Learning Coordinators, the actual maximum crept up to 24. The second was that the geographic areas that the Learning Coordinators in this province were expected to cover were much larger than in the other provinces. Rather than having to travel a few kilometers to reach their schools, many of them had to travel many miles. The third was that there had been a teachers' strike in this province that closed the schools for two months unexpectedly. The fourth was that the vans and motorcycles for Learning Coordinators still had not arrived. This was a problem also in the other provinces, but its effects in this province were especially severe because
of the greater distances to be traveled. The fifth was that the Learning Coordinators had to devote some of their visits to the collection of the achievement, attitude, and demographic data specified in the assessment design. This cut even further into the time available for constructive supervision.

The combined effects of these problems are illustrated in practical terms in the following example:

One of the Learning Coordinators we interviewed is responsible for 22 schools, the closest of which is 23 miles and the furthest of which is 110 miles from his home. He has been on the job nearly 15 months. He was given three months of training. He started making his rounds last November. He has no conveyance. Because of this and a brief illness, he was able to visit only four of his schools in November. In December, he did better. He was able to get to 15 before the start of the winter vacation. Schools were closed during January and February for vacation; he planned to resume his rounds when they reopened on 3 March. But then came the teachers' strike, and the schools did not reopen in March or April. He could not resume his work till sometimes in May. But now he could not make supervisory visits. It was time to administer the achievement tests that the current evaluation protocol requires. This kept him busy into June, until the start of the summer vacation. When the schools once again reopened on 27 July, he still could not get back to teacher supervision. He had to begin collecting the demographic data that the current protocol also requires. This chore still occupies him today. In his case, 15 months of "expanded supervision" consisted of one supervisory visit to 19 of 22 schools.

Clearly, an evaluation design that interpreted the results in this province as an indication of the utility of expanded supervision (as would the existing design) would not do the concept justice.

At the other extreme, the province that provided the greatest amount of additional supervision had limited its caseloads to only 6 schools per Learning Coordinator, and had assigned an additional Supervisor for every 60 schools. Weekly visits were common. In the provinces between these extremes, patterns varied because of difficulties in recruiting individuals to fill all of the positions.
(10) Studies of the supervision component should be designed so that the outcomes at each site are related to the inputs actually provided at that site. Aggregating the results, even within provinces, seems risky.

The data that we collected on the provision of instructional materials illustrate another type of discrepancy that can occur in the conduct of evaluations. It results from the use of information collected for administrative purposes (in this case, information on procurement) as an indication that the inputs have been delivered (in this case, to the schools). There was no way to tell from the central files that the distribution of the materials that had been purchased had not been completed in two of the provinces because of transportation difficulties. Any assumptions about the status of these inputs based on the "official" data would have been wrong for many schools. As in the case of the supervision component, meaningful studies of the effects of these inputs also required school-by-school input information.

The utility of collecting information on inputs is not limited to the detection of discrepancies. Targets of opportunity also can come to light. In one province the instructional materials had been prepared by local professors rather than purchased, and took a number of different forms. Some of those we inspected were "programmed teaching" modules that try to structure what the teacher does in class. This is an approach that has been found to be highly successful in other countries. There was an opportunity to evaluate it for use in Pakistan even though it had not been included in the design of the project. Such opportunities clearly should be exploited.

(11) Because they could add appreciably to the design of effective Type II interventions, the use and effects of the instructional materials developed in this province should be given special attention.

Apparent Effects and Accomplishments. Establishing which of the outcomes envisioned in the Appraisal Report in fact was being achieved lay well beyond the scope of the reconnaissance. We set out merely to look for signs of progress, to help identify the effects that should be studied more closely. If there were signs that some of the early (intermediate) outcomes were being achieved, trying to pin down these effects over the next ten months was probably worthwhile. In the absence of such signs, the chances that an additional investment in studying an outcome would contribute useful information were slight.

In looking for early signs, a roadmap is important. The sort of roadmap that we used is shown in Figure 2. It portrays the expectations or "rationale" underlying the supervision component. The question was whether or not the specific changes hypothesized had started to occur.
To answer this question about the supervision component, we relied chiefly on the "critical incident technique," which has the twin virtues of being both open-ended and objective. We asked Learning Coordinators and Supervisors to

- Think of the last time that you had an opportunity to do something during the course of a school visit that you thought really helped.

When the respondent began to describe such an episode (or "incident"), we probed for sufficient factual detail to insure that the account was genuine. Then we asked for a second example.

The findings were impressive. We obtained reports of accomplishments related not only to the outcomes predicted in Figure 2, but also to numerous others. In addition to obtaining community contributions and support, raising teacher and pupil attendance, and conducting on-site training sessions, Learning Coordinators were taking actions to strengthen school management, representing the school's interests at the district level, negotiating for the school with contractors and landlord, pitching in to help restore access to schools following a flood, substituting for absent teachers, etc. They emerged from our data as a sort of educational analog of the agricultural extension agent, doing what had to be done whenever and wherever it needed done to hold the system together.

(12) The supervision component may turn out to be a "success story" of major proportions. Its contributions should be documented in full detail in the follow-on proposal.

Most of the incidents that we collected came from the two provinces that had the most generous supervision arrangements. Having a fairly small caseload of schools that the Learning Coordinator visits often and gets to know well may be an important prerequisite to the success of this type of intervention. Because of the cost implications, the trade-offs associated with various patterns of caseloads should be examined in depth.

(13) Cost-effectiveness studies of the supervision component should be included prominently in the plans for the next ten months. These studies should be carried out with site-by-site (not aggregate) data, for the reasons earlier noted.

The accomplishment cited in the critical incident data also provided information about the other components. There were 21 separate incidents, for instance, of villagers donating land, money, or labor in response to a Learning Coordinator's request for help in building or repairing a school, adding a boundary wall, or installing fans. For our modest sample, this was a sizeable number.
(14) Support for self-help projects may be stronger than had been thought at the time of the appraisal. These possibilities should be explored particularly with reference to the enormous shortage of classrooms.

The reports of the Learning Coordinators also spoke to the accomplishments of the in-service training component. The help that they reported they gave the teachers in designing and using teaching aids was based directly on the impetus to the use of teaching aids that had been provided by these courses. They had the desired effect.

(15) Short, two- to three-week training courses appear to be an effective vehicle for teaching specific skills, such as the use of teaching aids. An expanded program of short, narrowly focused courses should be considered for inclusion in the follow-on project.

The evidence on Assistant Teachers was sparse, but uniformly positive. In the samples of Assistant Teachers and certified teachers we interviewed, the former seemed to be the more alert. The Supervisors and Learning Coordinators who had observed Assistant Teachers at work agreed with this assessment. They also reported that the Assistant Teachers who had been recruited locally were especially effective in dealings with the community, as the staff appraisal had foreseen.

(16) A further investment in the evaluation of the Assistant Teacher component is indicated. The rationale appears to be sound.

The findings on the construction of teacher residences to attract qualified teachers were generally negative. Recruiting teachers to live at schools in remote locations had proved to be difficult. Most of the units were empty. Those that were occupied, moreover, seemed to be occupied mainly by teachers who would have lived in the community in any event. There was no evidence that the availability of residences had contributed materially to the recruitment of individuals who otherwise would not have been available. We comment on some of the possible reasons for this below.

Important Process Elements. There are two basic reasons for studying process elements in a program evaluation. One is to try to determine why a certain component is not producing the expected results, with a view to fixing it. The other is to identify the elements that are important to the success of the components that do seem to be working, to insure that these features are retained as the program is expanded. We collected data of both types.

On the construction of residences for female teachers, which appeared not to be working, we solicited hypotheses about the reasons for
failure from our female respondents. The explanation that most of them suggested was that the provision of furnished quarters is not enough for women expected to live in remote areas. Adequate security (in the form of watchmen) also must be provided. So must basic amenities, such as electricity and water. They pointed out that efforts to recruit teachers from other locations had been successful only when the teacher had a relative or companion willing to live with her. To make the jobs attractive to a larger sample, there would have to be additional inducements.

In the studies of the next ten months, more detailed and reliable information could be developed on these points. But doing so did not seem to us to be cost-effective. Given that the concept of recruiting Assistant Teachers in remote locations not only seemed to be working but also offered the additional advantage of building closer links with the community, it was not necessary to look for ways to make the residential program effective.

(17) The collection of additional data on the construction of residences for female teachers is not likely to contribute to the proposal. The resources available for evaluative studies should be devoted to other components.

The process data that we collected on the supervision component, which did appear to be working, focused on possible threats. One that jumped out at us from the project documentation was the friction that was likely to develop between the established inspectorate (i.e., the Assistant Education Officers) and the individuals in these new positions. No attempt had been made to rationalize the structure. The new positions had merely been grafted onto the existing arrangements at the project sites. This ambiguity might be tolerable for the duration of an experimental project. But, on a permanent basis, we were quite sure it could not last. By an overwhelming margin, our respondents confirmed this suspicion. At a number of locations, the relationships between the established and new supervisors were openly hostile.

Other administrative problems also came to light in the course of our discussions with the Supervisors and Learning Coordinators. None was content with the existing conditions of service; all felt underpaid. Part of the reason was that the new salary schedule that had been adopted since the project began leapfrogged the pay level of many primary school teachers above the pay level of the Learning Coordinators. Our respondents indicated that they would not continue to work under the existing arrangements; one announced her resignation during the course of our visit.

(18) The conditions of service for Supervisors and Learning Coordinators (and Assistant Teachers, who voiced similar complaints) should be specified in the proposal. Evidence should be presented that these will attract and retain the kinds of people desired.
Another type of process data that seems potentially important was mentioned earlier. This is the data available on different sizes of schools. Observations of the way in which facilities of various sizes typically are used, and the associated effects on teacher and pupil behavior could be extremely useful in planning construction programs.

Important contextual Factors. Finally, it is important also to look for factors outside the project that should be considered in developing follow-on plans. Some of these may be long-standing characteristics of the environment that have taken on new significance in the light of dynamics revealed by the project. Some may be the result of events that occurred since the project began. In multi-year projects in particular, new developments that affect the project are likely.

The major development that had occurred during the year prior to our reconnaissance was the decision to begin a program of at-mosque schools as an additional way of meeting needs at the primary level. The plan was to build classrooms contiguous to rural mosques for co-educational classes at Grades 1 to 3. The regular curriculum would be taught at these new facilities by the imam and a secular teacher. Positions analogous to those of the Supervisors and Learning Coordinators also would be created to insure effective supervision. Because of the prestige and influence of the mosque, strong community support for this type of school was expected.

In principle, this new initiative could complement and strengthen the follow-on project. Having two options to choose from in meeting the needs at different locations should permit each to be used maximum effect. But, to achieve this result, coordination was essential. The construction component of the follow-on project could not be designed efficiently without knowledge of the locations at which at-mosque classrooms will be built, for instance. Nor could the conditions of service for Supervisors and Learning Coordinators be considered viable if their counterparts in the mosque program were being paid at higher levels (as had been proposed).

(19) The relationships between the follow-on project and the program of mosque schools should be delineated over the next ten months, and described in detail in the comprehensive proposal.

Another recent development was a project that distributed "teaching kits" to primary school teacher under the auspices of a UN agency. In our interviews, these kits were mentioned frequently. Both the teachers and Learning Coordinators had found them useful in their work with teaching aids. This suggested that similar kits (and perhaps also programmed teaching modules of the type described earlier) should be added to the project.

(20) The inclusion in the follow-on project of (a) programmed teaching modules and other teaching aids that would be introduced to the teachers initially in (b) short, in-service courses, and then serve as the
basis for (c) on-the-job training by the Learning Coordinators may be a highly effective way of tying all three of these components together.

On the basis of the data that we collected, this combination of activities emerged as a potentially powerful (Type II) approach to the reform of teaching styles.

Among the long-standing aspects of the environment that took on additional significance on the basis of our findings were a number that seemed relevant to the problem of school consolidation. One was suggested by the sample of schools cited as the best in each caseload in our "nominations" approach. A disproportionate number were schools that were attached to middle schools (so that Grades 1-8 were taught at the same location). It would seem useful to explore the use of this approach on a larger scale. Another concerned the segregation of boys and girls at the primary level. In one province, we noted that 29 of the 40 "boys' schools" on which we collected data in fact had co-educational enrollments. This is another area in which the limits should be explored.

(21) Activities during the next ten months should focus not only on the "experimental" treatments that were introduced by the project, but also on deviations from standard practice in the regular educational system that may contribute useful ideas.

Implications for the Project in Pakistan

The action implications are embodied in the 21 numbered suggestions above. Each of them establishes a specific, potentially important link between the kinds of information that the project can provide and the kinds of practical answers it is intended to produce. Each operationalizes the diffuse plans and hopes that went into its design in terms of specific, goal-oriented steps that fit the project as it actually evolved. Implementing all of these suggestions probably is not feasible. But it should be possible to implement a sufficient number to produce a well documented proposal for the follow-on phase. Within the framework of the existing evaluation design, this could not be done.

The incremental utility of the reconnaissance in this project may have been somewhat greater than will be the case in other projects. The evaluation design was unusually diffuse; more sharply focused plans could have been developed from the first. But, even in tightly structured assessments, the payoff of a midstream review is likely to be substantial. Reformattting the initial plans in the light of what actually happened as the project unfolded is simply a good idea.

The incremental costs were extremely modest, moreover. The investment in the reconnaissance totaled less than six person-weeks.
evaluation design was unusually diffuse; more sharply focused plans could have been developed from the first. But, even in tightly structured assessments, the payoff of a midstream review is likely to be substantial. Reformatting the initial plans in the light of what actually happened as the project unfolded is simply a good idea.

The incremental costs were extremely modest, moreover. The investment in the reconnaissance totaled less than six person-weeks. Provisions for one or more reconnaissances may be generally useful in managing Type I projects. We conclude with some speculations about this and the other procedural lessons that this case study suggests.

Implications for Other Projects

Broadly applicable principles cannot be extracted from a single case study. Not until the experience in this project has been pooled with others will it be possible to draw reasonably confident conclusion about its more general implications for the design of development projects.

What can be done even at this stage, however, is to extract the lessons of hindsight. Knowing what we know now, how would we have designed the evaluation component of the Pakistan project six years ago? If we could turn back the clock, what aspects would we change? The answers are not likely to be entirely generalizable. But, at a minimum, they will suggest some options in the design of future projects that it cannot hurt to consider.

A first and fundamental change that we would make would be in management philosophy. The exploratory, evolutionary spirit that was the dominant theme of the appraisal rhetoric would be given tangible form in the plans for implementation. The approach to supervision, evaluation, and management would be based on the following logic:

It is not expected that all of the inputs provided by this project will prove to be successful. Indeed, it is to find out which elements are the most promising that the project is being done. It will be important to monitor progress closely and carefully throughout the life of the project, and make appropriate, timely adjustments. By de-emphasizing the elements that turn out to be the less promising, and shifting more of the effort to those that show the greatest potential, we shall converge to the mix of effective and cost-effective inputs that the project is intended to reveal.

The expectation of continuing change and improvement would be the cornerstone of the supervision and management plan.

It will be noted that this philosophy is the opposite of the philosophy of experiments, which calls for exquisite safeguards to prevent
changes in midstream. It promotes a value system in which an evaluation that points to needed modifications is regarded as a contribution, not an indictment. Making changes would mean that the project is working, not that someone had been remiss. Had the project managers and project evaluators begun working together to identify needed improvements five years ago rather than just recently, there is little doubt that the evolution of a strategy to guide the development of primary education in Pakistan for the next five years would be correspondingly further along.

A related advantage of this incremental approach is that it would have softened the transition to the follow-on phase. Had the improvements suggested by the experience been incorporated as they were identified, fewer changes would be required in moving to Phase 2. Making many changes at once is problematic. Slippage must be expected if and when the project continues.

A second change, which follows from this philosophy, would be the establishment of an administrative mechanism for monitoring progress and making the indicated improvements. Hindsight does not tell us what sorts of provisions would have worked the best. But the experience does suggest four important components:

1. The first is the introduction and use of record-keeping procedures at the field sites that generate a continuous (quantitative and qualitative) history of project-related activities and their apparent effects.

2. The second is the conduct of periodic (probably annual) reconnaissance visits of the type described above, to compile and clarify the archival data, and collect information that is not readily captured by record-keeping techniques.

3. The third is to hold formal (probably annual) meetings of project and Bank representatives and appropriate third parties to review the feedback on progress and problems, and deliberate next steps.

4. The fourth is to commission special (aperiodic) studies of issues that cannot be adequately explored with archival data or brief reconnaissance visits.

Each of these components is elaborated below.

Record-Keeping Procedures. To the greatest extent practicable, the inputs and their apparent effects would be tracked with logbooks, diaries, and other archival procedures. Recreating history seldom is as accurate or cost-effective as recording it as it occurs. Many of the phenomena discussed earlier in this paper could and would be monitored this way. Some examples are: the number of classrooms at locations at which new classrooms are being constructed, the dates of occupancy and vacancy of
residences for women teachers, the dates on which instructional materials are distributed to schools, the dates of visits to schools by each Learning Coordinator and accounts of what was done, the levels of and enrollments in classes assigned to Assistant Teachers, etc. Much of the information collected during the course of the reconnaissance through special interviews would have been available from local records; proportionately more of the team's time could have been spent probing the dynamics.

Measures of impact would be limited, at least initially, to the intermediate outcomes that can be expected to occur within a reasonable span time; e.g., within the next two years. The collection of baseline data at the beginning of the project would be limited similarly. Causal relationships cannot be imputed over a longer span; too many other factors intrude. As the project matured, and there were signs that the early outcomes were being achieved, other indicators and baseline measures would be added.

On-Site Reconnaissances. The experience with the archival data for which the project did make provision confirms much other experience: records maintained in the field and aggregated centrally seldom can be interpreted without clarifications that central staff cannot provide. A more cost-effective approach is to examine the records on-site and resolve ambiguities with the record-keepers. This would be prescribed as one of the functions of the reconnaissance visits. A second important function would be to review project-related developments with a sufficient sample of participants to identify and probe phenominal that may require special attention. The nature of the shortfalls in teacher performance, the reasons for vacancies in teacher residences, and the circumstances surrounding self-help construction projects are some of the examples discussed earlier of issues that are better explored with interviews than archival data.

Periodic Progress Reviews. In the two supervision missions encompassed by this case study, the collection of data on progress and problems and the discussion of their implications with the Project Director were carried out as part of the same visit. This limited the amount of time that was available for reflection and exploration of alternatives. Completing the reconnaissance and circulating the report perhaps a month in advance of a multilateral review would be more productive.

The appropriate parties to participate in formal (probably annual) reviews in the case of this project would be FIU and PIU officials, representatives of the institutions providing technical assistance (British Council and/or SCRE), Bank representatives, and other scholars or officials selected by the Project Director for their special knowledge of certain topics. The explicit goal of each review would be to identify opportunities for improvements. The ambience would be one in which changes are viewed as forward-moving and not pejorative, as was stressed above.

Special Studies. One of the objectives of each review would be to identify issues that should be “illuminated” with special studies. The selection of topics would be based on practical project needs, not intellectual curiosity. A clear “need to know” to solve a particular
problem or otherwise improve results would be required. The impact of the
number of classrooms on teacher and pupil behavior was one of the special
studies of this type that was suggested above.

A third major change would be in the allocation of
responsibilities among the various actors involved in the collection of
evaluative data.

- The position of senior measurement
  specialist that was added to the FIU last
  year would be included from the
  beginning. We would realize that so
  large and data-dependent a project as
  this one cannot be managed adequately
  with occasional inputs from local and
  foreign experts.

- The basic monitoring and record-keeping
  system would be standardized throughout
  all four provinces, and managed centrally
  by staff with appropriate skills. Only
  special studies would be delegated to
  Institutes of Education or others not
  part of the central management cadre.

- Technical assistance activities (i.e.,
  the SCRE role) could be concentrated on
  (1) the design of the record-keeping and
  related data collection systems, (2) the
  periodic inspection and upgrading of
  these systems, and (3) such additional
  consultation as might be required from
time to time on the design of special
  studies. The in-country workshops that
  were conducted and seemed to be useful
  would be retained.

Most of these procedural suggestions are being
implemented in the appraisal of a primary education
project in Nepal that has a number of similar
features. This will provide a further opportunity to
assess this more systematic, empirically guided
approach to the management of Type I development
projects.
SECTION III:

LESSONS FROM AN EVALUATION/MONITORING SYSTEM

FOR THE HAITIAN EDUCATION REFORM

Richard Sack, Consultant
Education and Training Department
World Bank
1.1 The role of evaluation and monitoring in development has, over the years, been a subject of some debate. This paper takes the position that the *raison d'etre* of evaluation and monitoring is to affect directly both policy and operational aspects of the program under consideration—in the present case the broad reform of primary education in Haiti. This assumption will serve to focus the paper, both in its descriptive and analytic parts. Implicitly or explicitly, we shall be asking questions such as: What does it take for the results of evaluation and monitoring to be used in decisions concerning policy, planning and operations? What are the structural and communicative arrangements that can contribute to this goal? Conversely, what are the impediments to this happening effectively? What general lessons can be learned from this specific case? From the assumption that use of results is central, through a story line that includes description and analysis, we will arrive at conclusions pertinent to the organization of an evaluation/monitoring system that should have general relevance.

1.2 The educational reform in Haiti is of interest to us at three levels: technical; administrative-institutional; and, social-political. The technical level concerns the pedagogical and didactic work involved in defining and preparing new curricula and teaching methods. The administrative-institutional level represents the organizations where this work is carried out, implemented, put into practice and financed, as well as the various arrangements (often unclear and poorly defined) that exist between these organizations. The social-political level represents the set of reactions engendered by these efforts in the general public and, more importantly, several subsets of it. The Bank directly finances the technical level of the reform and, by necessity and more indirectly, the institutions where this work is taking place. The relative success of this investment is largely dependent on how the reform is received and the nature of the reactions engendered at the social-political level. In the midst of all of this has been placed an evaluation/monitoring function charged with the task of following the results of the technical work (mainly curriculum development) as well as conducting broader evaluations of the progress of the reform.

1.3 Throughout the paper we will use the somewhat awkward expression "evaluation/monitoring". This formulation is an attempt to acknowledge a paradoxical state of confusion and distinction between the terms "monitoring" and "evaluation." The distinction will be commented upon several times in the following analyses in the relevant contexts. As a first approximation toward a distinction, we can think of monitoring as an activity that continually tracks the on-going progress of the event under consideration. Evaluation is more within the realm of periodic studies of the event. These two terms will take on operational meaning as the paper delves deeper into the activities of interest to us.
1.4 We will first take a summary view of the reform in its institutional and larger socio-political context and then describe the evaluation/monitoring system followed by an analysis of its operations and its impact on policy. Throughout this exercise we will be particularly interested in the uses to which the information generated by the evaluation/monitoring system are put.
II. A SUMMARY VIEW OF THE REFORM

2.1 The reform was conceived to make formal schooling more accessible to a vast majority of the population as well as to make it more relevant to the prevailing social, cultural and economic realities of the country. To accomplish this goal, the strategic approach of the reform took the following form:

i) introduction of a ten-year basic school broken into three cycles of four, three and three years to replace the traditional primary-secondary structure;

ii) totally new curricula that use Creole, the national language, as the language of instruction for most of the first cycle of four years; oral French is taught in the first two years and written French as of the third year;

iii) retraining of teachers in the new curricula and appropriate new pedagogical methods;

iv) the development of a cadre of pedagogical advisers (conseillers pedagogiques-CP) working out of regional pedagogical centers (centres régionaux pedagogiques-CPR) whose major task is to work directly with teachers, follow up on their work and advise them on how best to apply the new pedagogical aspects of the reform;

v) the development of the Institut Pedagogique National (IPN) which serves as the major institutional base for the reform and which organizes all reform-related curriculum development, training, monitoring, and pedagogical research and development; and

vi) the development and application of a monitoring system whose function is to collect, analyze and interpret information on the reform's implementation and results in order to improve the management of the reform's implementation and, especially, to provide feedback to those technicians working on the curricular and pedagogical aspects of the reform.

2.2 All of this is to occur in an educational system which (in 1982) enrolled 658,000 pupils in 3,221 primary schools and had 15,000 teachers. The situation is particularly critical in the rural areas. In 1979 about 80% of the population lived in rural areas where (for 1977/78) the net enrollment rate was 24% as compared to 82% for primary schools in the urban areas. Other comparisons between primary schools in the rural and urban areas are instructive. In 1977/78:

- enrollment in rural schools represented only about 48% of total primary school enrollment although rural primary school-age children represented 75 to 80% of the total;
90% of the primary school-age children not attending any school were from rural areas;

primary urban school enrollments increased by 102% during the period 1960-78 whereas the increase was only 87% in the rural schools; and

40% of recurrent expenditures on education went to urban primary education (1976/77) and only 24% to rural schools.

The internal efficiency of rural schools was also particularly low as witnessed by the fact that only 38% of the children entering these schools reached fourth grade, which was considered the minimum amount of schooling required for literacy (in French). Furthermore, in a school system oriented toward achieving the primary school certificate, only 1% of those entering school in the rural areas succeeded in obtaining the certificate. In the urban areas about 48% of the children who entered primary schools reached grade four. It should also be noted that only 47% of the pupils are enrolled in public primary schools; the others are in the private schools, mainly associated with a church or religious organization.

2.3 The reform was officially introduced in 1979. To date, about one-third of grade one classrooms (1100) are using the reform programs. This has been accomplished in spite of hesitant beginnings, strong opposition from some segments of the public and, even, a reluctant Ministry for two years. To say that a class is using the reform program means that:

- the teacher has been trained for the reform;
- the curricula materials are those produced for the reform;
- Creole is the language of instruction with gradual introduction of French; and
- there is regular supervision of the classes by the CPs.

The education profession has (often enthusiastically) accepted the concept and practice of the reform and is making substantial progress in putting it into place.

2.4 Those aspects of the reform related to curriculum and instruction implied major investment in all areas dealing with the organization of education as well as the teaching process itself. This investment (financed by the Bank) required important material, intellectual and organizational resources. The major areas where this investment was made were (and still are) in: curriculum design and development once the basic pedagogical objectives were determined; elaboration and experimentation of didactic materials; training of teachers and supervisory-level staff; and the organization and implementation of the monitoring system.
2.5 The implementation of the reform has two basic dimensions: pedagogical/didactic; and, institutional/administrative. These two dimensions, and the myriad activities that comprise them, are intimately intertwined. The core of the work in implementing the reform is the relatively technical pedagogical and didactic work involved in the definition and preparation of new curricula and teaching methods. Learning objectives must be defined; curricula are developed and tested in conjunction with the new teaching methods; textbooks are prepared and produced; training programs are developed and organized according to the available free time of the different teachers, inspectors, school directors, CPs, and so on.

2.6 None of this work would be possible without an infrastructure of institutional arrangements. In Haiti, the Institut Pedagogique National (IPN) is responsible for all the pedagogical and didactic work including the training activities. Most of these activities are financed by the Bank project, which is administered by a Project Implementation Unit (PIU) in the Ministry of Education. The PIU facilitates the transfer of the required funds to the IPN upon authorization of the Ministry and the Bank. Teachers are appointed by the Ministry, which also controls the inspectors. Furthermore, the actual rate of implementation depends on the Ministry and its attitude towards the reform which, in turn, is related to the prevailing public and political climates concerning the reform.

2.7 The aspect of the public's attitude toward the reform and the predominant attitudes of the political class is particularly crucial and difficult to assess. Given the nature and number of interventions of varying types needed to carry out the reform, all it takes to create problems, delays, bottlenecks, etc., are a few people in relatively sensitive or important places who are lukewarm, at best, to the idea of the reform and its implementation. Therefore, the management of public opinion and the attitudes of the political class are of particular importance to the degree and rate of implementation of the reform.

2.8 This analysis suggests that a systems perspective of the reform could be demonstrated by a set of three concentric circles where each outer circle represents the environment in which its respective inner circle is operating as well as providing the resources needed by this inner circle to function. For the case at hand we have the pedagogical core (where the technical resources needed to implement the reform reside) which receives administrative and material resources from the surrounding web of institutional and administrative arrangements. These institutions, in turn, (and the individuals working there) receive their political resources (i.e., legitimacy) from the prevailing attitudes concerning the reform. Figure I shows a representation of the system.
Figure 1: A Systems View of the Haitian Educational Reform
2.9 The provision of resources is not without a cost and the consumers of the resources (the inner circles) must be able to demonstrate the utility of the efforts furnished (by the outer circles). In practical terms, this means that the pedagogical core is responsible for improving the learning results of the schools, increasing their internal efficiency and, especially, providing effective literacy to all children who complete the first four years of basic schooling. This is one place where an evaluation/monitoring system should come into play in two distinct ways: (i) to demonstrate the pedagogical results of the reform to those institutional authorities that provide the resources and sanction necessary to implement the reform; and (ii) within IPN, to provide the feedback needed to rectify and improve on the original pedagogical and didactic work. In other words, an evaluation/monitoring function should operate within the inner circle, between the inner and the middle circles and between the middle and outermost circles. This means that the information produced by an evaluation/monitoring system will flow in several directions: (i) within the inner circle where it will serve as feedback to those people working on the pedagogical and didactic aspects of the reform; (ii) from the inner to the middle circle in order to gain the administrative and material resources needed to carry out the pedagogical and didactic work; and (iii) from the middle to the outer circle in order to gain the legitimacy and political resources needed to follow through and expand the reform efforts. In the next chapter we will take a close look at this, especially with respect to the information needed for each interface.

2.10 The network of institutional support in the inner and middle circles is accountable (however informally) to the political authorities and the general public. The history of the educational reform in Haiti has indicated that progress in its implementation is particularly dependent on some form of acquiescence (however passive) by the general public and that the government is likely to retreat in the face of strong and well articulated opposition to the reform. This means that there must be some demonstration of results that are comprehensible to those who make-up the outermost circle, and this demonstration should be partially the fruit of the work of the evaluation/monitoring system. In the short run, increased internal efficiency and better pedagogical results would constitute such a demonstration. In the long run, however, there would be a need to demonstrate that the reform is yielding better external efficiency.

2.11 As systems models go, this is basically a political one. This should be of no surprise because of the very nature of an educational reform, let alone that of an educational system itself. In Haiti, as in most other developing and developed countries, education has long been perceived as the surest road to social mobility and any modifications of that road are bound to raise questions and suspicions. In this context, a reform of the existing system requires considerable resources, cooperation, coordination and good will on the part of myriad actors playing a wide variety of roles (technical, administrative, political, with some roles passive and others active). By its very nature the play is difficult to prepare, the audience arrives with an a priori critical attitude and the execution of the roles is highly complicated and complex. We can see, therefore, the need to bring onto the stage an evaluation/monitoring system that will provide feedback to some of the actors, information to others and, with luck and work well planned and executed, comfort to the audience.
III. THE EVALUATION/MONITORING SYSTEM

3.1 Evaluation and monitoring activities were written into the Appraisal Report of the Third Education Project. The Project Appraisal foresaw these taking place at the IPN and in the Ministry of Education's (MEN) planning unit (which has an evaluation section). The Appraisal Report stated that the IPN would undertake monitoring work focusing on the productivity of key educational inputs, whereas the planning unit of the MEN would do evaluation work in the area of teacher reactions to improved career prospects (which have yet to be implemented) and the reaction of the private schools to the reform.

A. Roles and Functions

3.2 Now that we have put the evaluation/monitoring system onto center stage our task becomes one of describing and analyzing its roles, functions, organizations, operations, uses and problems. To start with we shall look at the roles that it is expected to play and the functions it is supposed to serve in the overall implementation of the reform. As we can see from the way the stage has already been set, the evaluation/monitoring system for the Haitian educational reform is playing at two levels and in two dimensions. Each level represents a different audience as we have already seen from Figure 1.

3.3 At one level we find an audience composed of educational professionals, technicians and decision-makers within the educational system -- i.e., the immediate clientele. These are mainly the people within the inner circle and they need the type of information that will guide them in their work and decision-making. For example, the technicians working at the IPN on curriculum development need feedback on how the experimental programs have worked, the types of problems encountered, where particular difficulties lie, etc. Also, the teacher training programs can be continually improved as a function of information on particular problems the teachers encounter in the classroom and with their teaching guides. In other words, the pedagogical results of the evaluation/monitoring system are of particular interest to those people (mainly at the IPN) involved in planning and developing the down-to-earth aspects of the reform that deal mostly with the actual teaching and learning that goes on in the classrooms. These people also need information on the needs for retrained teachers in order to plan training activities; they also need to know how many reform classrooms are in operation and just where they are in order to plan the distribution of didactic materials and various follow-up activities undertaken by the CPs.

3.4 At this level we are talking about monitoring activities that track the results of on-going activities and provide feedback to front-line "implementators" of the reform. This information is generated within the inner circle of Figure 1 and, by and large, remains there to be used. This includes information on learning results, teacher effectiveness and the effectiveness and availability of teaching and learning materials. That information pertinent to the administration and organization of the reform (e.g., number of reform classes, teacher placements, logistics, etc.) is of particular interest to the middle circle which is more concerned with aspects of administrative and material support.
3.5 At the other level we have an audience composed of the general public and policy-makers and decision-makers working within the political and administrative systems whose opinions and actions can influence the educational system and the reform. This would include people in the outer and middle circles who are the top people in the MEN as well as those who have the ear of the President, who is the ultimate source of authority in the country. These people generally try to be in touch with the various expressions of public sentiment concerning the reform and experience indicates that they will seriously take into account any strongly held opinions that could contribute to political waves. However, what type of information can the evaluation/monitoring system provide to these people? Results, and "answers" to the following questions: Do the children (especially in the rural areas) learn better with the reform program? Is literacy easily and lasting acquired? Does the use of Creole actually facilitate the learning of French and just how well is French actually learned? Is this major change in the educational system going smoothly and are its costs within reason? Are the leavers and graduates from the reformed schools able to find work and adapt to the conditions of the labor market? Are four years in the reformed school (i.e., completion of the first cycle) sufficient to be able to adapt to the demands of the labor market, read and write and "make-do", especially in comparison to the problems of drop-outs from the old school system? Much of this information is supposed to be furnished by the evaluation/monitoring system put into place. From the IPN there should come the monitoring information concerning pedagogical results and the numbers of pupils in reform classes. The planning unit of the MEN is supposed to be able to provide evaluative information on non-pedagogical questions. Of course, some of these questions will not find answers until some years have elapsed—especially questions concerning the external efficiency of the reform.

3.6 The two dimensions of the evaluation/monitoring system refer to different types of information with each dimension executed by a different agency. The IPN collects information on the pedagogical/didactic dimension whereas the planning unit of the MEN is supposed to collect information and perform studies on a dimension that deals with the external aspects of the learning processes themselves such as the administrative comportment of relevant institutions, the behavior of the labor market to the reform and matters of costs and finances. Each dimension, therefore, serves a different function in the operation and implementation of the reform as well as informing different actors on the overall stage of the reform.

3.7 We now have a clearer perspective on the distinction between evaluation and monitoring—a perspective that is operationally relevant. Basically, each corresponds to a different level. The monitoring activities are happening at (and for) the level of those working most directly with the reform and within the major implementation agency for the reform which is the IPN. This is the inner circle in Figure 1. Some of the information produced by these activities is also of use to the planners and administrators in the middle circle and would include information on the rate of overall implementation and the administrative and material bottlenecks discovered in implementation. Some of the monitoring information could also filter to the outer circle, especially information
on pedagogical results (this, however, has yet to happen). Evaluation is more closely related to the middle circle with the information being generated at several places (the planning unit of the MEN, the PIU and sometimes at the IPN). This information is conceived mainly for the planning and policy making that takes place in the middle circle. This information is likely to filter to the outer level, especially as there is a need for political and financial resources in a climate of doubt (concerning the effectiveness of the reform) and competition (for financial resources).

3.8 The scene is now set for an overall view of the stage upon which the evaluation/monitoring system is operating. The diagram below (Figure 2) gives this view in a two by two format—two dimensions for two levels (or, audiences). Within the cells we find the particular functions being played by the evaluation/monitoring system on each part of the stage.
### Figure 2: Dimensions and Levels of an Evaluation/Monitoring System

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pedagogical-Didactic (within the schools)</td>
<td>A</td>
<td>Feedback for adjustments to initial programs and the design of new curricula; definition of training needs for teachers; understanding of factors contributing to pedagogical results; internal efficiency.</td>
</tr>
<tr>
<td>Administrative-Institutional (external to the schools)</td>
<td>C</td>
<td>Information on administrative blockages, bottle-necks and procedures; information on the relevance of curricula to the world of work.</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>General information on reform's results; information that could sway opinion if the information is strongly positive or negative.</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>External efficiency; administrative bottle-necks that can be corrected by political and policy decisions.</td>
</tr>
</tbody>
</table>

- Educational professionals, administrators and technicians. - Public opinion, political decision makers and educational policy makers.
3.9 The stage we are contemplating in the diagram is not an ideal one but purported to be real and operating. The pedagogical/didactic dimension exists in the form of the evaluation section at the IPN which is functioning, however imperfectly. The other dimension is less well defined in substance although its place of residence has been assigned to the Planning Department of the MEN where there is a subsection concerned with evaluation. And, in any case, the two audiences exist and are easily definable in terms of role and person.

3.10 We can now relate the operations of the evaluation/monitoring system as shown in Figure 2 to the systems model of the reform we have in Figure 1. This means mapping the cells of Figure 2 onto the circles of Figure 1. We have the following:

<table>
<thead>
<tr>
<th>Figure 2</th>
<th>Figure 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>cell A</td>
<td>inner circle</td>
</tr>
<tr>
<td>- B</td>
<td>outer circle</td>
</tr>
<tr>
<td>- C</td>
<td>middle circle</td>
</tr>
<tr>
<td>- D</td>
<td>middle and outer circles</td>
</tr>
</tbody>
</table>

3.11 The basic role of the evaluation/monitoring system in all of this is to provide feedback to relevant members of the audiences who are involved in some aspect of educational development and/or policy making. In other words, the evaluation/monitoring system provides information on a variety of relevant issues and matters to selected individuals who are in a position to act. This means that the distinction between audience and stage becomes blurred and that at some point the people at both levels are all in the same theater—the one in which the educational reform is playing. And, in many respects, we can see the evaluation/monitoring system as that force which enables these different audiences to play an informed and intelligent role in this living theatre which is the reform.

B. Organization and Structure

3.12 Evaluation and monitoring activities are supposed to take place in three distinct places: within the evaluation section of the IPN; at the Project Implementation Unit (PIU); and within the planning section of the Ministry of Education (MEN).

3.13 The Institut Pedagogical National (IPN). In many ways the IPN can be thought of as the major contracting agency for the implementation of the reform. Indeed, most of the Bank project financing for the reform is spent on activities executed by the IPN, including training activities, curriculum development, the production and distribution of textbooks for experimental classes, and the activities of the evaluation section. It is here, in the evaluation section, where we find the monitoring activities of the pedagogical and didactic work undertaken by the other sections of the IPN.
Before we get into the workings of the evaluation section we need to say something on the overall organization of the IPN and how the evaluation section fits into it. In addition to the evaluation section there are sections for training, field supervision, textbook production and curriculum development. The leadership of the IPN is composed of a director and a director of studies. Between them they are in charge of the overall operation of the IPN, the coordination of activities and the resolution of disputes between sections.

Curriculum development is broken down into subject-matter teams (written and spoken French; Creole grammar; writing and reading of Creole; math; social, physical and natural sciences; physical education; and pre-school learning) of pertinence, for the time being, to grades 1 to 7. The curriculum development teams produce both teachers guides and student textbooks. These teachers guides and textbooks are, to a large extent, the object of the specific monitoring work of the evaluation section. The production of guides and textbooks for each grade occupies a several year experimental cycle.

The pedagogical results in those classes where the experimental work is going on are monitored by the evaluation section during the experimental years (this will be more closely examined in the section on methodology and mechanics). The results of this monitoring are fed back to the curriculum development teams which are supposed to revise their guides and texts on the basis of the lessons learned from the monitoring. After three years of monitoring there is a full revision of the guides and texts (which can take 3-4 years) which are then produced in final form. The evaluation section monitors any one grade for only the three years of its experimentation with the new programs and curricula.

The evaluation section is (at present) composed of ten technicians who fill twenty defined functions, such as: data gathering and analysis; instrument design; collaboration and contact with each one of the curriculum development teams (i.e., one member of the evaluation section working continuously with each team); contact with the CPRs; and monitoring of the distribution of books. In order to cover the various areas of the section's work, the members of the group fill more than one position. These people are all university graduates (several have master degrees), who have, for the most part, been recruited directly out of university programs in the social sciences and education.

In order for the work of this section to be effective, extensive coordination of two types is essential: with the regional CPRs who administer the instruments in the experimental classrooms and are supposed to do a good deal of the observations and follow-up work; and, with the curriculum development teams who are supposed to be the prime beneficiaries of the monitoring work and incorporate its results into curricular revisions. The people working on curriculum development and as CPRs have been trained as teachers and have mostly worked as such. Their level of education has rarely gone beyond secondary school and those who have gone to university did so in various fields of education.
3.19 Within the IPN, the evaluation section appears (i.e., to the others) to have a rather curious status—it does not participate directly in the production of resources used by the reform such as curricula, didactic and pedagogical materials and trained teachers, nor is it involved in the operation of reform-related activities such as the CPRs. Yet, the evaluation section has the same organic relationship to the management of the IPN as have the other sections, meaning that its activities must be approved by the director, whose role and exercise of authority are critical to the functioning of the evaluation section in the larger context of the IPN.

3.20 The Project Implementation Unit (PIU). The primary function of the PIU is to administer the disbursements of the IDA credits for education projects being implemented. The second and third projects included the construction of basic schools and teacher training schools. All the basic schools built with project financing use the reform programs. In order to ensure that these schools function properly—i.e., receive the needed pedagogical materials, have qualified teachers who have received the required training, etc.—an education section was set up within the PIU. This section is headed by an inspector. When first created, this section's main responsibility was to advise the PIU on the selection of school sites. Now that the schools are built, the section intervenes mainly in the material, administrative and pedagogical aspects of the project schools. The monitoring of pedagogical results in these schools is the task of the IPN with the PIU working mainly with the inspectors.

3.21 This section of the PIU can also undertake periodic evaluative or quasi-evaluative exercises. For example, it is about to look into the matter of the educational radio program which is funded under the third project and whose activities and results are presently giving serious cause for concern. The PIU will undertake a survey designed to estimate the quality and extent of the use of educational radio. The results of this enquiry will contribute to a decision concerning the continuation of funding for the educational radio. This section is headed by a senior inspector and staffed by two young university graduates who have no specific training in evaluation.

3.22 The Ministry of Education (MEN). The Planning Department of the MEN includes a section that should do evaluative studies on various aspects of the school system including: the number of reform classes; issues of administrative arrangements; the effects of working conditions (salary, in-service training) for teachers on their performance; and the reaction of the private sector to the reform. To date, this section exists mainly on paper and its concrete activities are difficult to appraise. Some of the activities we would expect to find here are actually being done at the IPN. Although not what we would expect, this situation is the product of an accident of institutional development and is probably best left as is.
C. Methodology and Mechanics

3.23 Under this rubric we shall have to concentrate on the evaluative-monitoring activities of the IPN for the simple reason that there is no systematic work going on at the PIU nor at the MEN's planning department. The closest the latter has come to doing any evaluative work is a proposal for an evaluation of the administrative aspects of the educational system and a compilation of statistics on the state of reform implementation as of June 1983. Actually, the major product of the planning department is the statistical yearbook which is rich in descriptive statistical information without a hint of analysis. The wherewithall is there for useful analytic work of evaluative interest but this capacity is not being used for lack of conceptual and technical expertise (even though they have a functioning mini-computer).

3.24 With the exception of the rare occasional study, the PIU's education section is not involved in any evaluative work. It provides administrative and some pedagogical support to the project financed schools but does not collect information about educational processes that is (or could be) used for evaluative purposes. Basically, this section of the PIU operates like a school inspection office for the project schools, all the while working closely with the regional inspection offices.

3.25 This leaves us with the evaluation section of the IPN where there are on-going evaluation and monitoring activities using fairly elaborate methodologies and requiring good organizational and logistic support. The duties of the section are mainly two:

(i) Studies to examine the extent to which certain policy variables (teacher training, didactic materials and the physical conditions of the schools) have an effect on the pedagogical results of the pupils. This is done on a grade by grade basis.

(ii) The monitoring of the experimentation of new curricula material in a small number of experimental classes. The results of this monitoring are used by the curriculum teams for revision of the material.

In addition to these two major tasks the section is supposed to undertake periodic studies on a limited scale on issues of pertinence including the attitudes of parents and educators concerning the use of Creole in the schools; the influence of extra-educational factors (family, socio-economic status, etc.) on the educational outcomes of the pupils; and, an evaluation of the effectiveness of the distribution of didactic and pedagogical materials to the schools. The section also does an inventory of reform classes in the beginning of the school year.

3.26 The Evaluation of Pedagogical Results. After having established the number (and location) of reform classes, a 10% sample is selected for each grade being monitored (e.g., there are now about 1,100 grade one classes). Classes belonging to the sample are put into three categories:
(a) those which are visited once a quarter. These make-up about 80% of the sample and are geographically representative;

(b) those which are monitored on a semi-intensive basis by being visited once a week. They constitute about 13% of the sample; and

(c) those monitored on an intensive basis (4 times per week). They are the remainder of the sample and are all near the capital.

In all of these classes extensive information is gathered on the pupils and the teachers as well as on the school itself. Information is gathered by questionnaires and interviews as well as by observational techniques (especially for those classes which are closely monitored) and includes data on: the state of the school building; number of classes, pupils, etc.; the educational and socio-economic backgrounds of the teachers; the age, sex and background of the pupils as well as data on cognitive tests; and, behavior of teachers and pupils in the classes under intensive observation. These data are collected by CPs who have been trained especially for the task. The data are analyzed at the IPN and compiled into reports which are distributed within the IPN and sent to the PIU. There is no formal mechanism whereby they are communicated to the Ministry. An inventory of results is presented below (para. 3.31).

3.27 Experimentation of New Curricula. This is accomplished by an intensive monitoring of a small number of classes (five grade one; three grade three and three grade four classes at present) in different regions of the country. The instruments are the same as those used for the pedagogical evaluation and the classes are visited either every day (grade one) or twice a week. Again, it is the CPs who do the actual field work. The data are collected and analyzed by the IPN evaluation section.

3.28 The mechanics of the work of the evaluation section are fairly demanding, if not complex. Within the IPN, the evaluation section is supposed to maintain working relations with the curricular teams whose work they are monitoring. In practical terms this means that most of the members of the evaluation section are also working members of a curriculum team without having been recruited for that purpose (rather, they are appointed to the evaluation section and then delegated to work with the curriculum team). And then there is the matter of mobility and the section's ability to coordinate the field work of the CPs. This requires access to vehicles for both the members of the section and the CPs, credits for fuel and travel expenses as well as the capacity to pay the expenses of the CPs including travel to schools where the monitoring is taking place.

D. Results

3.29 There are two types of results that can be expected from evaluative and monitoring activities: the specific results of the investigations and the use to which they are put. This latter is the core of what interests us—i.e., the practical or policy-related effects of the technical results of the methodologies used. This will be treated later in a discussion of the impact of the evaluative and monitoring results.
3.30 Before we present the (technical) results of the IPN evaluation section's efforts it is necessary to mention that they are not nearly as extensive as the above presentation would suggest. There has been one major problem that has prevented a very large proportion of their work from being realized. The problem is lack of access to the means necessary to function (cars, gas, etc.) and the inability (largely that of the IPN's administration) to respect commitments toward the CPs who are the field agents doing much of the daily monitoring work. To this must be added the fact that the reform itself has progressed rather slowly in the past two years and there have been important delays in opening the reform classes as scheduled.

3.31 Despite these problems, there have been some results. Examples (with the year of the evaluation indicated in parentheses) are the following:

(a) The number of classes applying the reform programs (i.e., a class using the new curricula and related books, taught by a trained teacher and using Creole) is smaller than thought at the beginning of the school years. Some classes have reverted to old programs for lack of adequate supervision and materials. There has been, however, a steady increase in the number of these classes from one year to the next (1983).

(b) Many of the CPs are not sufficiently well trained for some aspects of their work and/or lack motivation (1982).

(c) There is a serious lack of coordination between the inspectors and the CPs (1982).

(d) Only about 50% of the rural classrooms are satisfactory in terms of area, lighting and outside noise (1982).

(e) Although the "legal" entrance age into grade one is six years, the average entrance age for the rural schools is 7.5 years and for urban schools it is 6.8 years (1982).

(f) Average class size is 50 in both rural and urban schools but the former have a standard deviation of 19 whereas it is only 5 for the latter.

(g) Classes are often up to a year behind the official program (1982). Grades 2 and 3 were judged unsatisfactory (1982).

(h) The results in Oral French at the end of grade two are satisfactory (1982).

(i) Math scores are higher for the religious schools than for the public schools (1982).
E. Discussion

3.32 As we have already shown, the major problem with the work described above is that, for reasons more administrative than financial and more political than technical, the means to carry out the monitoring and evaluation activities were not available. For that reason much of the programmed activity was not accomplished, especially as of the school year 1982/83. For these reasons the lack of sustained results cannot be attributed to an unrealistic or over-ambitious work program and evaluative design.

3.33 One question that arises concerns the design of the monitoring system and how to deal with the issue of comparison of results between the reform and the old system (still in operation as the reform is being introduced gradually). Another question is the issue of learning objectives and standards against which evaluative judgments could be made. The issue of whether or not to compare on a controlled basis the pedagogical results of the reform classes with those of non-reform ("traditional" is the Haitian term) classes enters into the center of the distinction between evaluation and monitoring. In the present case we are talking about a monitoring system that is keeping track of progress and problems in a system that may be improved (according to the monitoring results) but is the product of a well deliberated and (presumably) irreversible choice. To compare results in the two systems (reform and traditional) would be of real scientific interest but would not make great policy sense. The basic choices have been made—language of instruction, pedagogical methods, school structure, etc.—and the task we are discussing deals with those decisions as a given and then investigates the extent to which they are being fulfilled.

3.34 However, this requires some notion of objectives or standards against which decisions can be made concerning the changes and modifications needed in different aspects of the reform program. For example, the evaluation section concludes that the learning of written Creole is unsatisfactory. Compared to what? i.e., to what standard of satisfaction or to what learning objectives? There needs to be some basis upon which these judgments can be made, and this is absent from the work of the evaluation section. This absence is understandable given the novelty of the experiences being monitored. However, if learning is judged to be unsatisfactory, what does that mean: That the programs are in need of revision? That the goals are too ambitious? That there is a poor fit between the learning goals, the teaching materials and the teachers themselves? The need for objectives does not mean that they are immutable. Indeed, the results of the monitoring could very likely lead to revision of the learning objectives themselves as well as that of the programs and curricula.
3.35 This refers us back to the first part of this chapter where we presented a two-by-two picture of the roles of an evaluation/monitoring system (Figure 2). We can now see that only the role described in the upper-left cell A has been addressed. In other words, only the monitoring activities are being performed and the results of this work are almost exclusively used within the IPN. The roles described in the other cells are of greater concern to evaluation work and this has not been done, mainly for material and organizational reasons. And, of course, this state of affairs severely limits the impact that such evaluation activities could have on policy.

3.36 This also means that the flow of information one would expect to be generated by an evaluation/monitoring system is seriously throttled. Although formally transmitted to the Ministry (middle circles), the information effectively remains within the innermost circle of Figure 1. This leads us to conclude that for information to flow between the circles the roles described in cells B, C and D must be fulfilled.
IV. THE IMPACT OF THE EVALUATION/MONITORING SYSTEM ON POLICY

4.1 Now we arrive at center stage and the heart of the matter, because, the most salient output of any evaluation/monitoring system is the information it generates and how (and whether) it is used by policymakers and decision makers at different levels. In the present case, we are talking about information that has been generated formally by the mechanisms described above. This information is transmitted—formally and informally—to several sources which we shall associate with the circles in Figure 1: (i) the inner circle composed of the curriculum development teams which are supposed to use this information as feedback to be used in the revision of their curricula and the IPN team concerned with the training of teachers; (ii) the middle circle composed of the educational community at large which includes inspectors, private and public school directors and Ministry people all of whom are, collectively, in a position to have a large impact on the present and future implementation of the reform; and, more informally, (iii) to the outer circle which is the general public and political decision makers whose attitudes will determine the fate of the reform. Once the information from the evaluation/monitoring activities is generated and placed on the stage, it is up to these people to act (or ignore). However, we have just seen that although information is formally transmitted to the Ministry it is little used by them and effective use of this information remains within the IPN. The impact, therefore, of the information produced is dependent on the types of information produced as per Figure 2.

A. Intra-institutional Impact

4.2 Let us start out with those actors most immediately concerned: the curriculum development teams and the training section at the IPN which are supposed to use the monitoring information to revise their curricula. First of all, it must be noted that for material reasons the monitoring results were not always available when they would have been most needed by the teams. Some teams (math and oral French) did use the available results for curricula revisions. However, the team working on Creole reading and writing has simply refused to take the results into consideration, questioned their validity and, even, the competence of the evaluation section to evaluate their work. Also, the training section generally ignores the work of the evaluation section, claiming that it is not specifically pertinent to their activities. The nature and strength of these reactions raise several questions.
4.3 It appears that although the curriculum teams theoretically accept the monitoring work of the evaluation section (with the exception of the Creole team which does not even accept it), they have a difficult time putting up with the results in practice. Each team does its own work and considers that it, alone, is competent to judge the results. Also, they claim that poor results are not necessarily their fault—they could be the fault of the teachers, or the method of monitoring, or the poor physical conditions of the school, etc. The fact that each curriculum team works alone means that members are not predisposed to conceive of their work as existing in a set of pedagogical, didactic and environmental conditions that must be taken into account.

4.4 This is related to the way in which the technicians of the evaluation section were recruited. They were all recruited from outside of the IPN to serve in the evaluation section and then delegated to the various curriculum teams. As noted earlier, they are generally better educated than their colleagues in the curriculum teams and are unlikely have much experience as teachers or administrators. These delegated evaluators are not always considered as bona fide members of the curriculum teams and, for four of seven of these teams, communication is poor between the evaluation section and the teams. If the evaluation section had been constituted by members of the different curriculum teams plus several specialists for matters of design, methodology and analysis, the recalcitrance of the curriculum teams might well be reduced. For one, they would be less likely to contest the competence of the evaluation section. Secondly, the informal communications would be facilitated. This, however, would be a viable approach only when there is a fair degree of stability in the personnel, which has not been consistently the case at the IPN.

4.5 This leads to another—and perhaps major—problem in the receptivity and willingness of the curriculum teams to accept the monitoring results: fear which is associated with a misunderstanding of its role and function. Haiti is a society where there is little or no tradition of job security, where there is a practice of frequent changes at all levels on the basis of a change at the top and where authority is vested more in relatively particularistic norms (e.g., membership in a clan or group, friendship, and loyalty) than in more universalistic norms (e.g., scientific or technical competence, proven experience). In such a setting (which is certainly not unique to Haiti), it is normal for people to feel particularly vulnerable and sensitive to anything that may look like criticism of their work which, they suspect, may one day be used against them when there is a change at the head of the institution. In other words, the institutional norms that we implicitly associate with effective evaluation and monitoring and, indeed, are a prerequisite to effective monitoring, are not in place in Haiti. What has been conceived in a context of relatively universalistic norms has been more or less transposed to a society where the prevalent norms are relatively particularistic. In this context, we see that what are designed as monitoring activities are subjectively perceived as evaluative in the judgemental sense and this confusion is detrimental to the effectiveness of the monitoring function.
4.6 Another source of conflict between the evaluation section and the others is related to the fact that the former are more highly educated than the others and have a greater tendency to see themselves as professionals possessing technical skills. And, of course, they are doing the "evaluating" and presumably, can feel immune from the (real or perceived) pressures and sensitivities felt by the members of the curriculum teams. In other words, there is a perception that the evaluation section is an elite—they are better educated, are not recruited from the mainstream of the education profession and are one step removed from the front lines of reform implementation. Also, with a higher level of education (generally overseas), there comes a more developed sense of professional identity which also tends to be associated with the development of relatively universalistic norms. This difference makes it all the more difficult for the members of the evaluation section to comprehend the reticences and reserves of the members of the curriculum teams. This situation would also argue for the modification of the recruitment policy into the evaluation section.

4.7 We are thus faced with an overall context which is structurally conflictual. The two parties to this conflict are left to work it out among themselves, with the contrary being often the case. One of the reasons for this lingering state of conflict and limbo is that there is no effective exercise of hierarchical authority that exerts leadership and decides in case of conflict. Not that this is impossible in the present organizational structure of the IPN; rather there is no tradition of such leadership—i.e., the IPN's management does not comprehend the nature and extent of the problem. Why? Perhaps because of a lack of understanding of the basic political nature of any evaluation/monitoring enterprise in an environment not habituated to the normative (attitudes) and structural (job security) contexts necessary for such an undertaking. And why would that be the case? Perhaps (in part) because there is no tradition of evaluation/monitoring.

4.8 This state of affairs raises another question: are there built-in structural or organizational mechanisms that could remedy or, at least, palliate these deficiencies? The present analysis suggests (for the case at hand) that we might look to changes of leadership structures and recruitment policies. Another way to mitigate these tensions would be to establish formal procedures for "litigation" when conflicts arise between the evaluation section and other teams. Such a set of organizational arrangements could involve the exercise of leadership in a spirit of arbitration.

B. Extra-institutional Impact

4.9 Outside of the IPN little heed is paid to the results of the evaluation section. For one, their work is conceived in such a way as to be consumed mainly by the other sections of the IPN—i.e., their work is more in the realm of monitoring than evaluation and the former is of little direct concern to policymakers and opinion leaders. The potential is
there, but for the time being it is mainly negative. This is the case mainly because the generalization of the reform is still largely a matter of political will in a political climate where the most articulate voices (i.e., in terms of political resources) are reticent or hostile to the reform (they are mostly the urban bourgeoisie). There is, however, an increasing acceptance of the reform within the educational community which appears to be approaching a point of coalescence as more and more educational leaders throw their support to the reform (e.g., vocal support by the heads of the more prestigious Catholic schools and the adoption of the reform programs by the Methodist schools). If unambiguously negative results were to be produced by the evaluation section they might well be taken up actively by the opponents of the reform. This, however, has not happened and one reason could be because there is no direct comparison of short-term results between the reform and traditional classes. Strongly positive results could also have an impact. However, the design of the monitoring system plus the lack of communication between the evaluation section and policy makers outside of the IPN make this unlikely. We see, therefore, that the impact of the results on the middle and outer circles in Figure 1 is largely a function of the prevailing political climate.

4.10 This climate, however, has been slowly evolving to a position more favorable to the reform. There are two reasons for this: (i) the increasing support from the educational community; and (ii) decreasing resistance from various parts of the population once the initial shock of the reform has passed and as impressionistic evidence favorable to the reform accumulates and is communicated informally. The irony of it all is that the formally structured evaluation/monitoring system is foreign to this evidence as well as to how it gets communicated.

4.11 Until now, the one set of results that has been taken into account by the policymakers is the census of reform classes. This information has been used simply to know the extent of the reform's implementation which, in turn, serves to know how much effort and investment will be necessary for generalization, if the potential will is there.
V. CONCLUSION

5.1 The most striking conclusion that derives from this analysis is that the evaluation/monitoring activity under consideration is highly political. This is the case even when viewed from the narrowest perspective which is that of the intra-institutional context within which it is operating, presumably, at a rather technical level. We say political because it generates conflict within that institution—the IPN. This conflict arises from the specific normative and structural contexts in which the evaluation/monitoring is taking place which is in an environment little habituated to disinterested criticism formulated for purely corrective and meliorative ends. In other words, rare are the Haitians (working in the IPN, at least) who feels sufficiently serene and secure in their posts to realize that a monitoring-feedback-correction process is not an evaluation of their work. This is further complicated by the lack of a tradition of monitoring and evaluation, meaning that people tend to confuse monitoring results with criticisms.

5.2 This leads us to conclude that when designing an evaluation/monitoring system, special consideration has to be given to issues of management. Whereas methodological issues are relatively straightforward, management issues are more subtle and as complex as the relations between the actors on the stage where the evaluation/monitoring activities are playing. Also, this is further complicated by the likely possibility that many of the actors (e.g., those whose work is being evaluated or monitored) have little experience with the processes of evaluation/monitoring-feedback-correction and their behavior could take an unexpected and apparently irrational course once their work is the object of evaluation/monitoring. The management problem becomes one of how to encourage acceptance of the information produced by the evaluation/monitoring activities. Our analysis suggests the possible usefulness of introducing structural palliatives (since a real solution would involve changes in attitudes and patterns of job security in Haiti) within the existing institutional arrangements.

5.3 In the case of the IPN, resolution could (or, could have) come in the form of: (i) a different policy of recruitment of technicians into the evaluation section (it is too late for that now); and, (ii) an effective exercise of leadership conceived to settle disputes between the evaluation section and those destined to use its results. The matter of recruitment refers to the structural integration into the monitoring processes of representatives of those very groups whose work is being monitored. Basically, this would be a specifically appropriate variation on the theme of participatory evaluation which becomes particularly pertinent in settings where evaluation and monitoring are incongruent with the prevailing normative conditions (i.e., relatively particularistic). The other, and more immediately practical, aspect of the palliative (qua solution) would be the exercise of leadership in the form of binding arbitration. In this case, the management of the IPN would have to arbitrate disputes between the evaluation section and the curriculum teams. However, in order to accomplish this role effectively management would have to be able to demonstrate technical competence and have been following the work of the different sections.
5.4 Another issue that arises is that of who initiates a piece of evaluation or monitoring work. Much of the monitoring work is a matter of routine and, as such, should be generally accepted by all concerned parties. However, when it comes to non-routine evaluative studies (which the evaluation section undertakes from time to time) it is likely that they would have greater impact if they were to be initiated from outside of the section—i.e., if the perception of the utility of the study came from a policy oriented sector. There is always the danger, and probability, that studies initiated by the evaluation section will not be in the realm of the policy concerns of those making policy. If these latter do not actually initiate evaluative studies, they should at least be consulted in order to increase the possibility that the results would be used.

5.5 This brings us to the issue of the use and impact of evaluation work and how to increase the probability of its perceived applicability to policy making. One could conceive of evaluation and monitoring activities as tools at the disposal of the policy makers that become useful to them only when they know how to use and manipulate the tools. As long as these evaluation/monitoring tools are controlled and used mainly by specialists the probability is greater that they will be of minimal use to policy makers. This suggests that there is a need to engage the policy makers and the presumed consumers of evaluation/monitoring results in the definition of the work and, whenever possible, to have them involved in the work itself. A parallel approach would be more in the line of public relations whereby the evaluation/monitoring specialists are simply more outgoing and make special efforts towards their clientele of decision makers and in the present case, curriculum developers. For the model presented in Figure 1, this implies that there is a need to strengthen the communication of information from the inner to the middle circles and from the middle to the outer circles. This can best be effective when the outer circle is aware, interested and sensitized to the concerns of the middle circle, and likewise from the middle to the inner circles.

5.6 Although derived from a particular case, many of these conclusions could have general applicability. They point to the need to look carefully at the socio-political environment in which an evaluation/monitoring system is being placed—both from the micro and macro perspectives. It is particularly important to be aware of the sensitive reactions an evaluation/monitoring function can evoke in a setting which is little used to such activities. In these cases—which are common wherever evaluation and monitoring are undertaken—structural palliatives should, and can, be developed. Indeed, this should be of no surprise to anyone who has done evaluation work anywhere as avoidance behaviors are a frequent response to evaluative efforts. And, from such phenomena have developed strategies of participatory evaluation or, at least, the involvement of the clientele in the conception and execution of evaluative and monitoring functions.
VI. LESSONS

6.1 It is necessary to develop a clear understanding of the political context in which an educational reform is operating. This context includes the education profession, the administrative apparatus, and the larger context of opinion leaders and the public opinion. This requires an analysis of support and opposition as well as the relative strength and articulateness of these positions.

6.2 From this one should derive hypotheses concerning the uses to which evaluation and monitoring results will be put (and by whom), as well as the sensitivity of these activities.

6.3 One can now assess the political and cultural climates into which an evaluation/monitoring structure is placed and give special attention to the management of evaluation/monitoring in the overall context. Special attention needs to be paid to how the results may be received by people who may feel their work is being criticized. This requires an assessment of the nature of (perceived and real) job security and stability.

6.4 Particular attention should be paid to policies governing recruitment into an evaluation/monitoring section. If the personnel in this section are all very different in background (education, experience within the education profession, professional loyalties) from those whose work they examine there can be a divorce in perceptions and communications between the evaluation section and those for whom the results are intended.

6.5 It is wise to assume that evaluation/monitoring activities will engender conflict. A structure for resolution needs to be established in advance that would include exercise of leadership and arbitration that is acceptable to both parties.

6.6 Some form of participatory evaluation should be considered. This could take the form of recruitment into an evaluation/monitoring section of individuals from sections whose output is the object of the evaluation/monitoring. It could occur by having the consumers of the evaluation/monitoring results define or commission the work to be done.

6.7 Evaluators (groups or sections) should be fully aware that the purpose of their work is that it be used in operations and policy making. This awareness should lead them to something of a marketing strategy for their services—i.e., the establishment of communicative devices and links that would increase the probability that the evaluation/monitoring information produced would be used by the relevant "clientele" of policy makers and operations people.

6.8 From a methodological perspective, a clear decision needs to be made concerning the standards against which evaluative judgments are made. If there is no control group, the cognitive and behavioral standards should be made explicit.
Now what? How can all of this analysis help the situation in Haiti? In other words what are the design implications of this analysis now that the project is at an advanced stage? For the case at hand, we have to accept a large measure of the institutional arrangements already in place. However, there are areas where modifications could be made and new structures could be created. Some ideas in these directions are outlined below and presented by institutional relevance.

1. **IPN.** One has to accept the fact of the evaluation section's methodology and staffing, neither of which pose egregious problems. The problem that needs to be addressed is that of the relations between the evaluation section and the curriculum team. This requires the establishment of procedures for appeal, arbitration and decision enforcement. The following sequence could apply:

   a) If the curriculum team is not willing to incorporate the monitoring results into their work this should be made known to the director (or his delegated authority) of the IPN.

   b) The director would listen to the arguments of both parties and come to a decision concerning the work of the curriculum team, (i.e., whether and how to deal with the monitoring results).

   c) This decision could have an impact on the work of the evaluation section in terms of its methodology. In this case, any modification would be made in consultation with the relevant team.

   Whatever procedures are developed, they should be formalized within the IPN; this means recognition that monitoring can engender tension and conflict, and mechanisms for resolution should exist.

2. **Project Implementation Unit (PIU).** The question here is how can the PIU monitor the investments in the reform. Since the IPN is the major executing agency for the reform and, as such, is financed by the project, a good part of the PIU's monitoring work would involve the IPN. The monitoring staff of the PIU is two to three persons.

   We have already shown that implementation of the reform is largely dependent on the political will of the government. This means that if the PIU is to monitor the progress of the investment in reform it must operate at two levels: the technical, which includes the IPN and the schools; and, the political, which is the Ministry. Monitoring activities would mean keeping abreast of the Ministry's intentions (allocation of funds; directives concerning new reform classes) as well as keeping track of the IPN's work in training and curriculum development. The work with the Ministry would involve informal contact, information gathering and assessment of intentions. The work with the IPN would be more formal in nature and involve a process of accountability of the IPN's work.
In practice, the monitoring of the reform would consist of informal activities without a specific methodology. However, this does not exclude periodic studies to assess the effectiveness of specific subcomponents of the project (such as educational radio), especially when there is an issue of new or continuing financing. This monitoring work could include the following activities:

a) periodic audits of the production of pedagogical and didactic materials;

b) keeping track of the quantitative accomplishments of the training section (how many teachers trained; what are the needs);

c) keeping informed of the work and results of the evaluation section;

d) assessing the impact of the monitoring results on the work of the other sections; and

e) periodic visits to the CPR's to assess their work in the field.

The conclusions drawn from these observations would be communicated to the director of the IPN, the Minister and the Bank. Also, it should be clearly understood that these conclusions would be taken into account during the process of appraisal of the IPN's yearly work programs.

3. The Ministry of Education (MEN)

a) The ministry has just established a committee on reform implementation that could play a central role in the use of evaluation and monitoring results in the processes of reform implementation. The following recommendations appear appropriate for this committee's work.

i) It should be in continuous contact with the IPN's evaluation section and regularly receive its findings. This would be ensured by appointing the head of the evaluation section as a member of the committee who would be required to inform the committee of the findings.

ii) The committee should approve any periodic evaluation study undertaken by the IPN. It would be preferable if the committee actually commissioned such studies as they appear helpful in the light of implementation problems.

iii) The director of the PIU should also be on this committee. This would facilitate understanding of the problems and procedures involved in financing reform implementation.
iv) the committee should maintain contact with the planning department of the MEN and define its relevance to the processes of reform implementation. This could include the commissioning of studies and specific analyses of existing data.

b) The planning department has a sub-section concerned with evaluation. To date, no work of any interest to reform implementation has been produced by them, and present staffing makes such production unlikely. However, the planning department has produced large volumes of data (the statistical year books) and possesses a mini-computer. This combination could enable the planning department to perform valuable analyses and projections. This would require an additional staff member capable of doing statistical analyses, projections and operating relevant software.

4. Overall coordination. The central problem in the integration of evaluation and monitoring results into the implementation of the reform is one of ensuring and coordinating the flow of relevant information to the decision and policy makers. This coordination could best be ensured by the Minister's reform implementation committee which would have to establish formal tentacles to the IPN, the PIU and the MEN's planning department. In operational terms this should imply membership on the committee by people from those organizations.
SECTION IV:

SOME QUESTIONS AND LESSONS FOR THE BANK

FROM THE THREE CASE STUDIES

John Raven, Consultant
Education and Training Department
World Bank
Introduction

The Role of the Evaluator

The Case for the "Illuminative" Approach

Barriers to Effective Evaluation

Arrangements for Policy Research

The Institutional Frameworks in which Policy Research Work

A Constituency for the Work

Staff Development and the Role of Consultants

Conclusion
Introduction

The three studies raise a number of important questions, and contain many lessons, for the Bank. Some of the questions and lessons have to do with the methodology of evaluation. Some have to do with the institutional frameworks in which evaluators work. Some have to do with the role of consultants.

One of the most fundamental set of questions was raised by Middleton: should the Bank be in the business of helping to initiate, and monitor the effects of, pilot schemes? Should it be in business of evaluation at all? If it should be involved in evaluation, should the Bank be primarily concerned with finding out, for its own purposes, whether funds made available to borrowers have been well spent (and should these schemes therefore be extended)? Or should it concern itself with helping to establish the policy development and monitoring structures - and local expertise - which are required in borrowing countries if their economic and social development is to be promoted? If indeed, establishing such structures is deemed important, the Bank would have to address all the questions raised by Sack: what kinds of evaluation and policy development activities, and targeted at what audiences, are required? And how, as Middleton leads us to ask, are these audiences going to be held accountable for acting or failing to act on the information they are given?

To promote discussion of these fundamental issues, we may set the questions which have been asked in the context of some more general observations about the growth of evaluation activity.

From the 1940's onwards, the importance of evaluation - and, in particular, the importance of initiating policy-development through a series of experimental pilot programs - has been increasingly recognised.

Unfortunately, although it has frequently been obvious why individual policies and programs (like the Pakistan Primary Education Project) should be evaluated, little attention has focused on the more general changes which have taken place in the way our society is organized and which have led to the increased emphasis placed on evaluation, policy development, and monitoring activities. It is these general changes which it is most important for the Bank to bear in mind when making arrangements to evaluate particular projects and programs.

It is possible, here, to introduce only the briefest discussion of these changes in the nature of our society; more detail will be found elsewhere (e.g. in Raven, 1984a).

Whereas, a quarter of a century ago, we still lived in an economy largely managed by the invisible hand of the economic marketplace, we now live in a managed economy and, increasingly, in a single, managed, world economy. The management of these economies is dependent on the generation and use of appropriate information. The structures which are created for the generation and use of this information are therefore critical to the economic and social development of the societies concerned.
Before moving on, it may be useful to draw attention to an important change in the way in which economies are managed. The economic marketplace provides people with frequent opportunities to vote with their pennies on many issues. To vote in this way, they do not have to make the reasons for their feelings explicit. In a modern economy, in contrast, management decisions are taken on the basis of explicit information, and the quality of those decisions is crucially dependent on the quality of the information available. Further, the control of cash flows is used to orchestrate the achievement of goals which have been decided through the political-bureaucratic process. That is, the goals are no longer established by the use of money. In other words, the role of money in managing the economy has been reversed. It is this fundamental change in the nature of the world economy which has led the Bank to try to evaluate the effects of its lending.

It may be useful to note that these arguments are not affected by the current fashion for "privatisation", since most privatised units will be subject to extensive, detailed, financial intervention, legal regulation, social monitoring, and social accounting. Such units, therefore, have little in common with institutions which came into being as a result of entrepreneurial flair and market forces.

The import of these remarks is this: the economic and social development of countries like Pakistan, Haiti, and the Philippines is crucially dependent on their creating policy-development and monitoring frameworks. It is therefore of the greatest importance that, when considering the funding, staffing and implementation of evaluation studies, the Bank consider the issues raised by Sack and Middleton - namely the funding and development of appropriate systems both to carry out the necessary research, development and evaluation activity and to ensure that attention is paid to the information so generated.

What then, can we learn from the case studies about the nature of evaluation, the institutional frameworks, and the personnel required to implement such activities? We will discuss each of these issues in turn.

The Role of the Evaluator

There are a number of important lessons about the role of the evaluator which deserve to be drawn out of the case studies.

One of these has to do with the distinction to be made between policy research and academic research. As Schwarz emphasises, it is incumbent on policy researchers, and, in particular, on evaluators, to use all available evidence to go well beyond their data in order to generate prescriptions for future activity. An evaluator charged with the task of assessing the merits of an experimental pilot project cannot reasonably complain that the inputs were not made at the right time to suit the evaluation design. The object of a pilot project is to build on, and extend, tentative insights and suggestions from earlier observations, thinking and research. It is to improve on the inputs by trying to make sure that they work, not to find out whether they work. It is to reject ideas which turn out to have been ill-founded. The fact that, as both
Middleton and Schwarz found, the inputs in pilot programmes are in a continual state of flux and development which differs from place to place and from time to time at the same place is therefore both inevitable and desirable. There can be no justification for not modifying a procedure which is demonstrably not working. What is more, as both Middleton and Schwarz also found, it is probable that most promising developments will have only weak effects in the short-term, because they will not be accompanied by necessary supportive changes in the system. An example of such supportive changes would be other teachers doing related things in order to build on and strengthen the effects of particular teaching techniques. To be fully effective, changes in educational activities may need to be supported by changes in peoples' perceptions of what society rewards, changes in parental attitudes, changes in the criteria applied to teachers by inspectors, and changes in the examination system so as to give teachers and students credit for having pursued alternative goals.

Not only will the effects often be weak, it is more than likely that it will not be possible to demonstrate any short-term effects on the educational outcomes which it is hoped to influence: the number of students exposed over a long enough period to appropriate educational experiences will be too small and the time scale will, in any case, usually be too truncated for the effects of cumulative experience to show up in outcome measures. Indeed, the time required to notice the need for, and develop, appropriate outcome measures may well exceed the time scale of the project. It is therefore frequently necessary to seek intermediate indices of the effects of the changes which have been introduced—to look, for example, at the effects of buildings on the uses to which they are put and not wait to measure the effect on teacher motivation and pupil attendance.

Instead of mounting outcome-focused studies, and complaining that variations from time to time and site to site upset their experimental designs, it is therefore incumbent on evaluators to generate evaluation designs which capitalise on that variance, and which lead them to ask, not only which variants work better, with what populations, and why, but also why the planned variations were not fully implemented at all, or certain, sites. Asking such questions may reveal critical deficiencies in the original understanding of the nature of the problem and the steps which were to be taken to solve it. Alternatively, such an analysis may pinpoint practical problems which may mean that it is necessary to abandon, or greatly modify, the original design.

It is equally unwarranted for an evaluator to complain that the number of students he is dealing with is too small to permit him to undertake a multi-variate analysis of the relative importance of the various inputs. A vast pilot project would cease to be a pilot experiment designed to assess the feasibility, and importance of, a possible program of activity. It would become a major activity in its own right.

As Schwarz notes, the objective of an evaluation of a pilot program is to assess the potential and practicality of a projected program or activity. It is to identify the potential long-term effects of properly developed inputs in the context of appropriate supportive changes. It is to identify the realism of that prescription. The evaluator cannot content
himself with showing that weak versions of the inputs, implemented in an inappropriate context, and probably assessed by poorly developed measures of outcomes, appear to have little effect. He has to go beyond his data and use his knowledge of the processes at work to indicate what the real effects of properly developed inputs are realistically likely to be when implemented in a more supportive context.

Schwarz's observation that the objective of an evaluation is to influence the future course of events carries with it another very important corollary: the evaluator's responsibility is to draw attention to all important outcomes of the activity. To fail to draw attention to an important outcome — whether desirable or undesirable — however "intangible", and however hard it may be to index by "objective" means — is serious. It will result in subsequent policy discussion — and therefore subsequent policy decisions — failing to consider important constraints and important benefits, or disadvantages, of the activity.

This emphasis on the importance of the comprehensiveness of evaluation is underlined from another perspective in Sack's paper. Sack argues, first, that the actual effects of any innovation will depend on what is done by administrators, text book writers, employers, pupils and parents as well as teachers. It is therefore essential to examine the role which each of "these — and other groups of people" play in determining the effects of the innovation — for the innovation may be re-interpreted in such a way as to negate its intent, or compensatory action may be taken to negate its effect.

But Sack goes further: he argues that since the objective of the evaluation is to ensure that the innovation is a success, the needs of all these groups must be addressed in any effective program of monitoring and evaluation designed to improve projects and programs. Teachers will need some form of credit for the extra work involved. Pupils will need tradeable evidence of the additional competencies they have developed. Parents will need evidence to reassure them that the time-honoured social functions of the educational system will continue to be performed — and not to the marked relative disadvantage of their own children. The general public (who vote the money or, even in a non-democratic system, can often bring effective pressure to bear to block activities they do not like) will need information on the external benefits of the system — such as whether it is contributing to economic and social development. Teachers and advisers will need feedback to help them to improve the educational processes which are being implemented. Administrators will need guidance concerning the most productive areas in which to invest further money, and the administrative procedures which will facilitate the innovation — provided they, like parents, have already been convinced that it is desirable.

The object of any proposed evaluation and monitoring activity must therefore be to provide each of these groups with different types of information. The evaluators' responsibility is therefore much wider than that of responding to the specific needs of their paymasters. It is to stimulate a whole process of public (or semi public) debate.
By now, we have found ourselves arguing that the evaluator's task is to provide a public commentary on the multiple issues involved. It is to comment on relevant issues which he had not been asked to comment on at all. It is to draw attention to drawbacks as well as benefits of the activity. It is to arrive at alternative options and scenarios for the future as well as to describe what was happening at the time and its effects.

Such a role is hard to reconcile with the most common understanding of the role of the evaluator - which is to answer administrators' questions. From our present vantage point, such a viewpoint would seem to be much too narrow. Schwarz's most outstanding contribution was, in fact, that he managed to change the questions which were asked and the issues which were being discussed. The view that the role of the evaluator is to structure the issues which get discussed - by a wide range of policy makers from politicians and administrators to parents and pupils - would seem to be much closer to the truth. But this is a role which, of necessity, most evaluators will fall far short of. We must therefore caution that it is not only important to beware of expectations which are too high in the sense of definitiveness and conclusiveness in the data collected, but also against wild expectations in relation to comprehensiveness - both in terms of the issues addressed and the audiences reached.

Sack's paper - and the question which Schwarz's paper led us to ask about whether the evaluator's role is primarily to answer questions - raises another topic of fundamental importance. This is whether the evaluator has an essentially different role in a socialized economy where centralised, communal decisions are taken or required, compared within a decentralised system where parents and pupils are explicitly encouraged to make decisions and choices.

My own view would be that there is no essential difference. In both cases, informed decisions based on information on the costs and consequences of alternatives have to be made. In both, some collective actions have to be taken to make individual decisions possible - educational options have to be provided and means of giving pupils credit for the competencies they have developed in the course of these alternative programs have to be available. In both, effective implementation is dependent on understanding of objectives and processes by a wide range of people - employers, administrators, advisers, teachers and parents. In short, the world-wide move toward privatisation and de-collectivisation of decisions does not, in fact, lessen the need for information - and the structures and organizations required to generate that information have to be provided collectively. There are, however, two important differences. If decentralised decision taking is to take place, the information generated by evaluators has necessarily to be public. Conversely, if communal decisions are to be taken, it is necessary to develop structures which will give teeth to the results of the information generated. This is Middleton's concern. In practice, the effective use of the results of evaluation requires that we pay more attention to both dissemination and implementation.
One final point which the case studies lead me to make explicit about the role of the evaluator is this: whereas the designs of the Pakistan Primary Education Project and the Philippine Radio Education Project assumed that the evaluator's task was to comment solely on educational issues - i.e. whether children "learned" better under certain conditions, or whether attendance went up - some of the most important lessons which emerged were systems lessons. This is clearest in Schwarz's paper, where some of the most important conclusions related to the building program or the role of learning coordinators. But it is true of Middleton's study as well, although he himself failed to note the fact precisely because of the way he defined his role. He attributed the fact that "no difference" was found between the experimental and control groups to the increased effort expended by the teachers in the control schools. Supposing this hypothesis were confirmed, what would the implications be? They would be that the introduction of quality control procedures which would give teachers and pupils credit for their achievements would probably lead to performance improvements which were at least as great as those which could be expected as a result of the introduction of new teaching techniques. This is a systems conclusion, not an educational conclusion, so it was overlooked.

Likewise, Schwarz's most important conclusions are not strictly related to educational outcomes: rather, they relate to a process which economists and administrators have all too frequently overlooked - it contributes to an examination of the links between the inputs (the hardware, the buildings, the staff, the textbooks) and the objectives which the provisions were intended to achieve. If the Bank is concerned that its investments yield the desired benefits - e.g. growth in personal competence - it will, in general, need to do much more to encourage the study of the actual workings of all the linkages between the financial inputs and the human resource and social development outcomes it hopes to achieve. Buildings can be built, teachers trained, textbooks written and printed, and examination-based evaluations mounted without its making the slightest difference to the development of the human resources so badly needed in Pakistan or elsewhere.

We may conclude this discussion by saying a little more about the distinction to be made between the role of the academic and the role of the evaluator. The task of the academic is to be as accurate and as theoretically relevant as possible. The task of the evaluator is to be as comprehensive as possible, to address as many as possible of the relevant issues, and to be as forward-looking as possible. It follows that sophisticated academics are likely to find the methods of the evaluator disagreeably crude and their willingness to make value judgements about other social processes unacceptable.

It also follows that an academic training is unlikely to be that which it is most appropriate for an evaluator to receive, and an academic environment is unlikely to be the most promising base for evaluation activity. It also follows that evaluators will need a career structure which does not require them to make "distinguished academic contributions" in order to obtain security of tenure. If evaluators are forced to work on soft money, and depend on academic recognition for their future livelihood,
it is probable that their work will be tailored to academic criteria and not
to the criteria which, as we have seen, it is appropriate to apply to the
work of the evaluator. Their work is therefore not only unlikely to answer
administrators' questions, it is unlikely to perform any of the other
activities we have identified. A separate, secure, career structure and
institutional base would therefore seem to be strongly indicated.

The Case for the "Illuminative" Approach

We have now explored the role of evaluation in a general way and
examined some of the tensions which exist between evaluation and academic
research. It is useful now to devote a few paragraphs to considering the
particular approach adopted by Schwarz.

We have already reviewed some of the constraints which led Schwarz
to abandon a classical stance. These included the fact that the inputs
could be expected to vary from place to place and from time to time
depending on resource constraints, local needs, and initial trials and
improvements. They included the fact that it would be impossible to
document the benefits using "hard" measures until long after decisions would
have had to be taken on whether to extend, modify or drop the project. Some
measures of their immediate effects and potential were therefore needed.

Schwarz noted that the effects of many of the inputs, such as
building hostels for women teachers in rural areas (in order to encourage
them to work there), were likely to take many years to show up in pupil
attendance and attainment. It was therefore inappropriate to use a classic
experimental design to assess their effect on such outcomes. Some other
measure of their effectiveness was required. This would not be an
educational outcome, but something which was, in the long run, likely to
lead to an educational outcome – such as the physical presence of teachers
or, even before that, use of hostels for appropriate purposes.

Another problem is that important effects of such things as
teacher training (itself dependent on the establishment of a suitably
trained network of learning coordinators) are likely to remain weak and
undetectable so long as understanding of the objectives of, and adoption of
the educational activities expected to reach, new educational outcomes has
not been widely diffused through the teaching profession. For a long time
the effect will be negated by teachers having to work in schools in which
pupil expectations are predominantly determined by the activities of more
traditional teachers, thereby undermining individual teachers' efforts to
change their behaviour. To detect the effect using a classical strategy it
would therefore be necessary to wait many years. Furthermore, if the
innovation is targeted at goals which were neglected by the traditional
educational system, it would be necessary to measure the effect using tests
which go beyond the outcomes commonly included in educational examinations
and evaluations, which few educational researchers would therefore know how
to construct, and which would therefore require several years of innovatory
research and development to develop.
Classical experimental and multi-variate designs also encounter serious problems stemming from the fact that they are generally able to cope with only one, two, or three, outcome measures at the same time. These might, as in Pakistan, consist of attainment, attendance, or retention. As we have seen, this is a critical flaw from the point of view of evaluation design, for the task of the evaluator is to document the effects of the processes he is studying on all important outcomes, whether or not he has "reliable" and "valid" measures of them.

Schwarz also comes close to noting that traditional attitude scales and traditional measures of academic attainment are not designed to yield feedback which will help teachers or pupils to improve their performance, which will help curriculum developers to improve the curriculum, or which will help textbook writers to improve their books. They are designed to discriminate between students rather than to discriminate between treatments. At best they can be used to "measure" the "effects" of "variation" in the inputs. Such "effects" tell us little about why an input was, or was not, working, or what should be done to improve it because the measures of both the inputs and the outcomes are too gross and undifferentiated.

Schwarz's position on this issue is, however, somewhat too extreme, partly because many administrators do believe that simplistic changes - like getting a couple of people to write a textbook - will have dramatic effects on pupil outcomes. The initial basis of Headstart was an example of such an over-simplistic idea. It is therefore necessary to retain measures of outcomes in order to find out whether the steps which have been taken are having the desired effect. It is no response to argue, post hoc, that no effect could have been expected because the inputs were so poorly implemented. It is important to find out whether the procedures as originally envisaged and funded work. It is precisely these "poorly thought through and developed" programs which would have been extended had not evaluators been around to question the wisdom of so doing and to highlight more "sophisticated" versions of the procedures which might actually work. (It is important, however, to recognise that getting any form of interpretable result requires careful design and also to recognise that the leap from a study designed to find out "whether it works" to one designed to find out "what works, under what circumstances, and how it can be improved" tends to be unexpectedly large).

There is another reason why large-scale survey type evaluation studies should not be jettisoned. They often contain - and experienced evaluators are adept at leading them to contain - items which can be used to identify potential ways of improving administrative procedures, curricula, textbooks and teacher training. Indeed, experienced evaluators are often skilled at using such studies, not for large scale multi-variate analyses of scale scores, but for fine-grained analysis of item statistics. Analysis of responses to individual items in tests can reveal a great deal about what is being taught effectively, and what might be done to improve curricula and teaching. Likewise, consideration of the answer given to individual questions in attitude and opinion surveys can lead directly to suggestions for improvements in administrative procedures, teacher training, relationships between teachers and pupils, and relationship between teachers and parents.
It can lead to an understanding of the barriers to the implementation of the procedures. The same comments apply to observations of behaviour. Careful, fine-grained, analysis of item statistics can be used to build up a useful picture of processes and outcomes, what is working in what ways, why not, and what can be done to improve it, and to raise questions about the relative importance to be attached to different types of outcome.

In view of these comments, it would be unfortunate if Schwarz's paper led the Bank to a wholesale rejection of large-scale attainment and attitude surveys.

Schwarz's own solution to the problems of evaluation discussed earlier, admirably illustrated in his paper, was to himself carry out a different kind of study - an "illuminative" study. He studied the inputs - the processes at work - with a view to discovering how to improve them, what the barriers to their effective operation might be, and what their long-term effects were likely to be.

In the light of what has been said about the role of the evaluator it is, however, important to draw attention to a number of limitations of Schwarz's contribution - for the issues on which he succeeded in focusing attention were precisely those which were easiest to measure. Such issues lie furthest from the heart of the educational process. The effect, if not the intent, of his work is therefore likely to be to reinforce researchers' already too dominant tendency to avoid even trying to measure the educational processes and outcomes which lie at the heart of student growth and development. Schwarz's approach therefore needs to be seen alongside, rather than taken to illustrate, the approach of those "illuminative" researchers who see their objective as being to study educational processes with a view to gaining some insight into what are likely to be their most important - if least tangible - outcomes. Their intent is precisely to focus attention on crucial educational issues which have previously been neglected. They argue that positivist researchers have much to answer for in focusing policy-makers' attention only on the low level outcomes and processes which are easiest to measure and failing to discuss the more intangible and harder to measure processes which cannot be reduced to variables bearing a monotonic relationship one with the other.

There is another lesson to be drawn out of Schwarz's paper. Schwarz, like the rest of us, clearly has most confidence in his conclusions when he has seen the processes he describes at work with his own eyes. The significance of this observation is this: it suggests that the primary task of the evaluator is to find ways of making intangible educational processes and outcomes - and organizational and managerial processes and outcomes - tangible and visible to all. This is an inventive and creative task which the Bank might be inclined to feel it was inappropriate to fund. Far from it: it is the most important activity to fund. When the intangible has been made visible it will not only be possible for evaluators to mount more sophisticated and convincing studies. It will also be possible to get more appropriate educational and administrative tools into the hands of practitioners - pupils, parents, teachers and administrators. This process of making the intangible measurable lies at the heart of the scientific process. Newton's and Ampere's most significant contributions, for example, lie in this area.
Barriers to Effective Evaluation

We have seen that the evaluation of pilot programs is crucial to the process of finding ways of deploying public funds more effectively, and we have reviewed a number of problems which have to be resolved before such work can be conducted more effectively. It may be useful now to explore why it is that so few evaluations are judged to be effective.

The belief that so few evaluation studies are worthwhile may itself need to be questioned, however. Evaluations are frequently deemed ineffective even though they have had a major impact on people's understanding of the nature of the problems and what should be done about them (Nisbet and Broadfoot, 1980). Thus, despite Middleton's evident disappointment with the evaluation itself, and the way in which the results were used, we find that the staff who undertook it have moved into influential positions from which they will undoubtedly share what they have learned. Likewise, it is inconceivable that the many people who have been involved in discussions concerned with the evaluation of the Pakistan Primary Project will not have used what they have learned.

Despite these observations, the question why inappropriate evaluations regularly get planned and executed needs to be taken seriously. Indeed, it is so serious that a joint committee made up of representatives of 12 organizations, including the American Educational Research Association and the American Psychological Association, was set up to consider the issue. Its report, *Standards for Evaluations of Educational Programmes, Projects and Materials*, was written by two dozen authors, commanded by four project officers, and supported by well over a dozen consultants. Yet the Joint Committees' conclusions were largely technological, and, even in that sphere, do not, in my judgment (Raven, 1982, 1985) come to terms with the problems involved in evaluating pilot programs which have been so ably identified by Schwarz.

We may first briefly review the logistic problems which prevent evaluations being conducted effectively. Those mentioned by Middleton included:

- lack of whole-hearted co-operation between agencies;
- temporary staff with insecure bases and career structures (This leads to the work being undertaken by personnel who do not know its background and who have little stake in the outcome, to time being spent on seeking other posts, to lack of commitment, to concern to meet criteria which would advance one's career, and to an undue concern with getting results which will please administrators);
governmental regulations preventing or delaying the availability of equipment (e.g. vehicles) and training;

changes of ministerial and Bank personnel, resulting in the outcomes of the evaluation being irrelevant to the concerns of those who are in post at the time they became available;

construction of unuseable buildings (in Middleton's case recording studies, in Schwarz's case, schools) and unsatisfactory equipment;

employment of attainment tests which are unlikely to register the effects of the particular input being evaluated;

over estimation, at the design stage, of the strength (and therefore detectability) of the effect to be produced;

belief on the part of the staff of the evaluation unit that they know sufficient about the methodology of evaluation to proceed without external assistance;

failure to specify the linkage between the outcomes to be achieved and the inputs to be made, the ancillary and supporting processes which were crucial to this change, and the ways in which the intermediate effects (which became inputs to the next phase of the process) are to be detected.

In Pakistan these problems were supplemented by others:

too few staff provided for in the budget, and failure to fill even those vacancies which existed;

red tape preventing the payment of those staff who were appointed;

inexperienced staff;

deployment of some of those who were appointed as evaluators on executive duties relating to the project inputs;

unwillingness of the base organization to release staff to work with the project;

a weak Federal Implementation Unit with little power, even if it wished, to command the compliance of personnel working to the provincial governments in the planned activities.
In planning evaluations, therefore, it cannot be assumed that officials from the governments and bureaucracies of the countries concerned will attach as much importance to evaluation activity as do officials in more developed economies, or that there are enough people with sufficient understanding of the logistic, personnel, and other resource requirements to be able to organize the studies locally and get them off the ground. It may therefore be necessary for the Bank to provide assistance even at this basic organizational and administrative level. While it may be thought that the best solution would be to start small and allow the countries concerned to develop their own expertise, the problems with which they are trying to grapple may well demand more urgent action.

The logistic and resource problems apparent in the Philippine and Pakistan projects, however, pale into insignificance compared with problems arising from the ways in which the studies were conceptualised.

Let us first look at the problems stemming from the fact that the studies were viewed as "experiments" intended to test the hypothesis that a combination of inputs identified in the course of previous projects would solve certain problems.

Given this orientation, it was assumed that the task of the evaluators was to find out whether the proposed changes had the desired effects. It would therefore not be necessary to collect data on which inputs were in place or on the educational programs which were in use: not only was it assumed that the planned provisions would be made at the right time - there was, at that stage, no perceived need to relate outcomes to inputs in order to find out what worked (let alone under what circumstances) or which inputs were unnecessary and could therefore be dropped. As in the evaluation of Headstart, this was an idea which was only introduced later - as it became apparent that no dramatic overall benefits would be demonstrable. The idea that it might be appropriate to conduct such a study could not be entertained earlier because, in Middleton's words "in an experiment there is little scope for mid-course corrections". The idea that one might set out to search for possible corrections and improvements was irreconcilable with such an orientation. Given this perspective, the Pakistanis could see no way in which information on the relative merits of different inputs (whether derived from a multi-variate analysis of data on outcomes, or from illuminative studies of the inputs) could be used. Indeed, it would be wrong to use it: the object of the exercise was to conduct a before/after study of the effectiveness of the total package. The proof of the pudding would be judged from the quality of the end product, not from an examination of the ingredients in the cake mix.

Schwarz's notion that the object of the exercise was to identify ways in which provision could be improved was not, in fact, something the Pakistanis themselves, or their advisers, could legitimately introduce: they were working to the Bank's design. (It had, in fact, been pointed out
on a number of occasions (World Bank, 1979) that the design which had been adopted would be unlikely to yield the necessary information). In actual fact, however, there were Pakistanis who understood the need to move away from the experimental design generated by the Bank. Why were they unable to move?

One reason was that, without Schwarz's model study to go by, they had no clear indication of the resource implications of alternative approaches or concrete information on the potential benefits of the approach.

Another reason had to do with the absence of research personnel. The initial proposals for the evaluation had made provision for substantial numbers of Pakistani research personnel (World Bank, 1979). At a later stage, the number was reduced by 80% but no change was made in the evaluation brief, or in the number of expatriot consultants.

This change conveyed a number of messages. Firstly, it gave the impression that the Bank thought that the evaluation was unimportant - how could a massive and important program of activity be undertaken by two evaluators in each province?

Secondly, it conveyed the impression that the Bank was more concerned with window-dressing than with obtaining useful results. How else could one explain the fact the elaborate design, and desired outcomes, were not accompanied by provision for the necessary personnel?

Thirdly, the disproportionate provision for expatriot consultants gave the impression that the Bank had little confidence in the professorial staffs of their Institutes for Educational Research and Teaching.

And, fourthly, the proportion of the budget earmarked for Western consultants smacked of the academic imperialism which has been identified in other aid budgets where the bulk of the provision, if not more, returns to the donor country. As Burgess and Pratt remark in their Polytechnics in Pakistan "it is not only more blessed to give than to receive, it is also cheaper" because not only do the funds provided by the donor institution return to the donating country, they are accompanied by the matching funding from the recipient!

Another difficulty in changing the approach has already been mentioned. The alternative to the experimental approach (which focuses on educational outcomes and, at best, attempts to trace back variance in outcomes to variation in the project's inputs using multi-variate analysis) focused on non-educational inputs such as buildings, materials, salaries, and working conditions. (The latter were, in any case, curiously ignored in the design). These were not likely to be of great interest to, or to promote the career advancement of, educationalists who worked in academic Institutes of Education and Research.

Another problem was the notion that the evaluator's task is to answer administrators' questions. No one in authority had asked the Pakistani evaluators to ask or answer the questions which it was Schwarz's most outstanding accomplishment to both ask and answer.
But there were other important processes at work too. One of these is that the results of enquiries along the lines undertaken by Schwarz were seen by the Pakistanis at the time as being likely to lead only to criticism of authority. Although new to Schwarz and the Bank, several of Schwarz's "discoveries" were common knowledge among those who were implementing the project and the evaluation. That, of course, is something of a truism, because one of the functions of the illuminative methodology adopted by Schwarz is to take otherwise "useless" knowledge and reformulate it so that it becomes useable both by those concerned and by others. The point is, however, that, in the absence of the rationale for the evaluation which has now been forcefully spelt out by Schwarz, and in the absence of sufficient distance from the project, the Pakistanis could see no way in which such information could be used to improve inputs. In the context of their understanding of the nature of the exercise in which they were engaged, it could only lead to criticism of the senior Bank and Pakistani officials who had formulated the project design. That, as Sack rightly emphasises, is something to avoid, and not only, as Sack would have us believe, by researchers working on soft money in societies governed by particularistic norms.

The "illuminative" material was not the only example of information which, absent a framework for using it, the Pakistanis declined to collect because it was likely to embarrass senior administrators. Other examples included information on teacher knowledge and understanding of educational processes, and information on the inputs which were actually in place. They avoided collecting both sets of data despite the fact that both were included in the original Bank design and despite the fact that, as Schwarz has again emphasised, they are crucial to explaining variance in the outcomes. The rationale for abandoning the collection of these data was partly the absence of the research staff needed to collect them, partly that proof of the pudding was to be found in the eating: if educational attainments went up, there was no need for measures of intervening processes. Nevertheless, there was evident relief that it would not be necessary to collect such potentially embarrassing data. (Such failure to collect or publicize potentially embarrassing data is, of course, far from unknown elsewhere! We are concerned here to understand the reasons for the behaviour, not to apportion blame for it).

Not only did they collectively decide not to collect such information (although some provincial teams [being accountable to their own governments and not to the federal implementation unit] went ahead and collected it anyway!) they did not draw attention to the implications (for both the rationale of the project and design of the evaluation) of statistical data.

* Other such knowledge included awareness that the foundation assumptions on which the project was built were untenable. It is highly likely that the "education" proffered will, given the occupational structure of Pakistan, not only not benefit the children concerned or the society in which they live (except in Sack's sense of enhancing the social mobility of some) but actually unfit them for life in either urban or rural Pakistan. There is widespread awareness of the effects of "over qualification" in India - a land of limited "employment". As was forcefully pointed out by a hotel porter, the problems of Pakistan do not derive from the lack of education among the rural population, but from the lack of management ability among the leaders and managers of Pakistan. (The same could be said, with equal force, of the U.K.).
which they and the Bank already possessed: the base-line demographic data for Baluchistan - highlighting, as it does, the small size of the schools, the distances between them and the fact that pupils have dropped out of schools well before the age grades at which the project was targeted - make nonsense of the project. Yet neither Bank staff nor the Pakistani researchers chose to draw attention to this fact (although Schwarz quickly rediscovered all from the personal observations he made).

The unwillingness of evaluators, especially those dependent on "soft" money, to come to conclusions critical of people on whose co-operation they are dependent if their work is to be continued, and especially of their paymasters, is well known everywhere. It is worse in societies governed by particularistic norms, and in those, like the U.K. and Pakistan, which discourage public supervision of public policy. Yet the effective operation of administered societies is crucially dependent on the publication and use of such information - and on related information contributed by whistle-blowers and others. Bank staff could usefully address their minds to the question of how social researchers are to be given both sufficient personal security and the right to publish and disseminate their conclusions to create a climate suited to the effective deployment of public - international - funds. It is one of the crucial questions of our time.

Another reason for failing to adopt an alternative design was that the researchers were academics, whose experience was of carrying out small-scale academic studies, rather than large-scale, policy-relevant, research. Their careers would also best be furthered by carrying out additional academic studies, rather than by obtaining useful results. (There are no academic kudos for an experienced professor of education in showing that a building is not used for the purposes for which it was intended; multiple discriminant analyses of test scores and "sophisticated" test construction procedures (however meaningless their results) would better serve that purpose!).

Then there was the (legitimate) perception by those concerned that they had best complete that upon which they had embarked rather than heed the advice of consultants. If they changed horses in mid-stream, in order to pursue the favorite topics of each consultant who came along, nothing at all would be accomplished.

But perhaps the most important of all reasons for failure to change the design was the fact the Pakistani researchers already "knew" how to construct tests and how to measure attitudes. They had experience of doing both things, even if the results sometimes left something to be desired by Western standards. This, in any case, was a defect which they felt they could remedy with a little advice and substantial resources. The fact that the Bank had chosen to provide mainly expensive expatriot advice and scant resources was mysterious. Many of the Pakistanis did not know how to use these techniques to conduct evaluations. They did not know - and nor did those who conducted much more expensive evaluations along the same lines at the same time in the United States - just how little studies carried out in this way would contribute to understanding the project's effectiveness or how to improve it. They did not know that all evaluations are always a mess, but with careful handling, something useful can usually be prised out of them - even if that meant continuing to analyse and mull over the
implications of the data long after the project had been deleted from the list of active research, and long after funding had run out. They did not know how to conduct "illuminative" studies, or what the benefits would be.

They did know that they would like someone to hold their hands as they went along and that they needed stability in the consultants who worked for them but they were unwilling to face the financial implications of that knowledge, for not only did each day of a consultant's time cost them as much as they would have had to pay one of their own professors in 6 months, they would have to make, and pay for, a continuing arrangement with the consulting organization.

Although these observations about the Project have even less to do with education than those made by Schwarz, they are of even greater importance to the Bank. For the Bank's overwhelming priority in preferential lending must be to establish the structures and expectations which will lead to its funds being well spent - even if spent on activities for which they were not intended.

What these observations point to is the need for the Bank to consider very carefully the institutional arrangements which it encourages its clients to make for policy evaluation, the links to be established between researchers and evaluators and the users of research, and the arrangements it encourages its clients to make with external consulting agencies.

**Arrangements for Policy Research**

We have now seen that:

(1) policy research and evaluation are central to the effective operation of modern economies.

(2) the process of policy evaluation does not mesh easily with traditional academic concerns and research.

(3) the whole process of policy evaluation is much more messy, and demands much higher levels of insight into social and educational processes, creativity and reflection than is commonly assumed. Policy research is not "normal science" in Kuhn's terminology.

(4) policy researchers need to target the results of their work at multiple audiences - indeed it is a mistake to think of Policy Makers as a single group of people, for policy is made at many levels.

(5) many classical prescriptions for curing the ills of policy research - such as Middleton's urging evaluators to focus sharply on "decisions" (such as whether radio should be used) and to adopt "simple" designs - are mistaken - because, to continue with the example, radio can be used for some purposes in some circumstances and not for other purposes in different circumstances.

(6) the classical experimental paradigm is a dangerous snare.
Because many of these conclusions run counter to current beliefs about the management and execution of policy research, we shall now consider the relationships to be established between policy researchers and the users of their research in more detail than did Sack when drawing attention to the problem.

The case studies and our reflections on them suggest that the Bank particularly consider the following issues when negotiating contracts for the evaluation of particular projects:

- The nature of the institutions in which personnel concerned with policy development, monitoring and evaluation will work, and the links established between those agencies and those who will implement the results of their work. Networks involving frequent consultations between several personnel from both sets of institutions rather than hierarchical or committee-based linkages might well result in more use being made of the results of research than was evident in Middleton's case study, and such networks of interaction are, in any case, indicated by broader discussions of the administration of public policy, such as those published by Schon (1971) and Raven (1984a).

- The constituency which will support policy evaluation and research and press for its continuity in the face of pressure to muzzle critical researchers. A broadly-based constituency of users—ranging from the public, through professionals, to senior administrators and politicians—is indicated from the accounts provided by both Middleton and Sack. Again multiple, frequent, and direct linkages between personnel at all levels, rather than hierarchical, representational, linkages seem to be indicated.

- The expectations which borrowers are encouraged to form, both of evaluators and social researchers and of their work. Schwarz's paper illustrates that the policy researcher's duty is to problematize the previously unproblematical, to ask and answer questions (rather than merely to answer administrators' questions), to identify issues which require the attention of policy makers, to stimulate and coordinate discussion, and to press for action on the basis of the insights gained. Many of these (such as raising questions and pressing for action) run counter to widely accepted views of the role of the researcher.

- The role of consultants in introducing the institutional arrangements, expectations, procedures and techniques which are required. This question is forcefully raised by Schwarz's paper, but it is an issue which merits much fuller discussion.
We will now discuss each of these issues more fully.

The Institutional Frameworks in which Policy Researchers Work. As is apparent from the case studies, it is far from ideal for policy researchers to be attached to academic or administrative units, particularly if they are able to devote only part of their time to their new role and if their career prospects remain tied to their old institutions. It is also unsatisfactory for them to be placed in the position of having to ingratiate themselves with politicians and administrators in order to be able to retain their livelihoods on "soft" money. A secure independent base with close links with both the academic community and the users of research is therefore indicated. Close links with the academic world are needed so that policy researchers can call on necessary theory, techniques, understandings, and knowledge of relationships between educational processes and outcomes, and so that they do not neglect to develop the fundamental theoretical insights to which applied social research invariably leads. They require close links with both administrators and the policy units' clients (for, as Sack shows, the users of research are usually not the people who pay for it) so that they can disseminate their research and identify previously unnoticed, but important and researchable, problems.

A wide range of personnel with different preoccupations, motivations and talents are therefore required in policy monitoring, evaluation and research units. In the case of education, it is necessary to have some people studying national problems; it is important to have some studying buildings; it is important to have some whose forte it is to make intangible educational processes and outcomes visible and measurable; it is necessary to have some who conduct large-scale, detailed and numerate, statistical studies; and it is necessary to have some whose forte it is to conduct illuminative studies.

But, whichever style of research is involved, it is important to remember that many of the most important benefits of the activity will come from the insights and understandings which the researchers have gained in the course of their work, rather than be immediately obvious from the "data" they have collected. Researchers therefore need opportunities to mull over, clarify, discuss, write up and disseminate these insights - and to explore their implications through discussion with administrators and other policymakers and users. Indeed, they need to be asked to research the "unmanageable", the "political", and the "intangible" problems which will be most likely to lead them to develop these insights. They need close links with policy researchers in other communities who are attempting to tackle similarly intractable problems. They need to be encouraged to investigate the constraints on policy makers' - including "users" - behaviour as much as the effects of the educational processes which are actually implemented.

Some of the most important personnel to be employed in policy research units will therefore be opinionated, outspoken, people interested in going well beyond their data in order to generate prescriptive advice. They will be passionately interested in their chosen topic and therefore open to the accusation of having vested interests and axes to grind. Such personnel will necessarily base their advice on assumptions about educational and social processes outside their area of research, and these assumptions may not be shared by others. It is important to remember,
however, that surfacing such assumptions is one of the most important contributions which social researchers can make - with the corollary that public debate will often force assumptions into the open. Personnel who have such strengths and commitments are unlikely to be prepared to drop one topic and take up another when the priorities of politicians and administrators change. One therefore cannot expect to recruit them on short-term contracts or expect them to be willing to change course in response to every change in administrators' priorities. What is needed is a rolling program which enables administrators to obtain quick advice on topical issues but which also allows the researchers to discern and pursue more long-term and important research questions. If it is difficult, in many countries, to envisage an appropriate niche, and pattern of relationships with administrators and the public for such personnel, it is even more difficult to conceive of ways in which external consultants can be recruited and deployed to develop the skills of local personnel who are interested in pursuing such a role.

To the extent that the role of the evaluator is, in part, to focus attention on new issues, to raise questions, and to provide useful information to a variety of audiences ranging from the general public, through teachers, to politicians and senior administrators, it is necessary to provide both time and facilities for numerous individual consultations, discussions, contributions to committees, and publication. It is also essential for a researcher to have close links with the users of his research so that he comes to take their problems seriously, reflect on ways in which these problems can be minimized, and undertake research designed to find ways of minimizing them.

Since it is new understanding which is most urgently required, it is desirable to have several units all hotly researching "the same" problem - for advance in understanding comes about most rapidly from public debate between researchers with different perspectives. Administrators' desire for tidy organizational structures therefore have to be held in check.

Likewise, as is evident from Schwarz's behaviour, if the results of research and evaluation are to be used, it is necessary to encourage researchers to campaign for action. One cannot, of course, generalize from a single case like this, but research by Roberts (1969) in the physical sciences, and Cherns (1970) in the social sciences, has shown that the results of research are rarely implemented unless the researcher takes it upon himself to force the issue. Yet the degree of commitment needed rests uneasily with widely-held views about the "neutrality" of the researcher and the academic.

The question of the mechanism to be adopted to promote the use of research - highlighted by Middleton's observation that "the project had not succeeded in its primary objective of supporting rational decision-taking" - deserves fuller discussion. It is a mistake to believe that there should be a one-to-one relationship between evaluation and policy-decisions - for decisions must reflect many more considerations than those with which the evaluator has been concerned. Nevertheless, evaluators do require some mechanism to help them ensure that the results of their work are fully debated and, if appropriate, used. Some mechanism is needed to give teeth
to the results of their enquiries. In other words, some mechanisms are required so that researchers can harness the energies of others (as in a feedback system) to press their case (perhaps through structures which involve the users of their research and the public in general). Likewise, procedures of accountability which will hold administrators accountable for implementing the results of research are also needed. This topic is crucial to the effective management of administered economies and will be discussed more fully below.

By now it will be apparent that the real problem is not so much to create an appropriate institutional framework in which, and base from which, social researchers are to work as to legitimize the creation of appropriate roles for them within the public service of modern society.

In some respects, therefore, our reflections point to the need to disperse social researchers around government departments. That would enable them to discern administrators' problems and help them to influence their decisions. But the need for mutual support and access to experienced fieldworkers, attitude-surveyors, test-constructers, coders, and data processing personnel precludes this. To compensate for the isolation in the special units which are required, it is therefore desirable to create special structures within these organizations, and networks of linkages with other organizations. "Flat" (non-hierarchical) units with numerous direct contacts between personnel at all levels in the administrative and executory hierarchy will be required. It is important to emphasize that, while the considerations mentioned above point toward the need for specialized policy monitoring and development units, a number of such units are required, some central and some local. This will help to stimulate badly needed debate about the implications of the research. The local units will also not only make it possible for more people to initiate research-based enquiries, they will also provide a mechanism whereby the results of research undertaken by other units (in the country concerned and in other parts of the world) can be incorporated into local decision-taking machinery. Care does, however, need to be taken to ensure that all these units are actively involved in research and not merely information-gathering and coordinating units which stockpile paper: they need proper links (of the sort mentioned above) with the users of research, and they need to provide opportunities for local personnel who have identified researchable problems to initiate such enquiries.

We may now consider in slightly more detail the steps which need to be taken to ensure that, when justified, the results of research are used. The "when justified" is introduced because much harm has come from the inappropriate implementation and over-generalization of the results of research. Examples include some of the extraordinary activities implemented in the name of "intelligence" (such as much special education), actions implemented as a result of studies of the correlates of academic achievement (such as restricted versions of minimum-competence testing), and the vast headstart program - which was based on Bloom's studies with 18 children. Such a tendency toward the use of social research to legitimate the implementation and continuation of practices which may actually need to be discontinued is perhaps discernable in Middleton's paper.
Nevertheless, the question is important. As Sack indicates, an appropriate framework involves links with parents, teachers, advisers and district education officers, as well as links with administrators and politicians. Many such links can only be implemented through publication, but personal contact has frequently been shown to be crucial to the diffusion of innovation.

But, beyond linkages, there is, as Middleton shows, a need for some mechanism to ensure that users actually consider the results of research. One way in which this could be achieved would be to require public servants to demonstrate in the course of formal accounting exercises that they have actually considered the results of evaluation exercises. Another possibility would be to create a mechanism whereby their work was open to public inspection in a way which would allow members of the public to ask whether potentially important results had been considered, what had been done about them, and why not. This would provide a mechanism whereby the various interested publics on the informed participation of whom the effective implementation of the policies will eventually be dependent would become better informed about the process. While, therefore, at first sight, the creation of such an open-government system would seem likely to delay the implementation of policy, it is probable that, in the long run, it would greatly increase its efficiency and effectiveness. Apart from anything else, it would reduce the level of suspicion and distrust which often surrounds the work of public servants.

We may now take a closer look at the question of the independence of researchers and their right to disseminate the results of their enquiries. As is clear from Sack's paper, it is important for policy researchers to be able to disseminate the results of their work to multiple users (policy makers) and to be able to do so without having to clear their publications through administrators who may have vested interests. This is not to say that policy researchers should be entitled to the last word on the subject concerned. On the contrary, Sack's analysis of the workings of public policy points to the need for open government in modern societies and to the need to legitimize public servants' ability to take a public stand on issues of public concern. Advance in understanding comes from public debate between groups investigating "the same" issues from different standpoints, not from "authoritative", definitive, enquiries. There is no such thing as a perfectly balanced and objective enquiry into issues of fundamental social importance. What interested parties need is not, therefore, the authority to suppress findings they do not like, but the right (1) to insist (if necessary) that the researcher include a brief disclaimer or objection (of one or two lines) from named people or institutions in his report and (2) the right to publish their objections and their own conclusions in detail elsewhere.

Finally, it seems appropriate to comment on the implications of what we have seen in the case studies, and said so far, for the contractual bases on which researchers work. It is clear that the most important benefit which society will get from the employment of social researchers is new insights, not answers to administrators' questions. Clearly, there is a great need for researchers to have scope to follow up questions they have uncovered. Equally clearly, there must be ways in which parents, teachers,
administrators and members of the general public can initiate enquiries into issues which they believe to be important and ensure that these issues are properly investigated. There is a clear need for mechanisms to mesh these interests and to establish priorities. Equally clearly, those mechanisms should not give undue weight to administrators' priorities. What an administrator needs in a researcher is someone who will take his problems seriously and address his mind to a proportion of them. He needs a mechanism which will enable him to get some answers to his questions now, and which will enable him to initiate activities which will yield answers to important questions in the future. While it is true that, as Schwarz notes, it is the researcher's duty to ask how an innovation will work after it has been further developed, it is also true that it is his duty to answer— as Schwarz did — questions which have not yet been raised.

The implications for the Bank and its borrowers are clear: they need to initiate enquiries, leave scope for their re-orientation, and, on top of that, to provide sufficient "fat" for researchers to follow up emergent questions and relate to the needs of users. The costs of such evaluation, development and dissemination units will still represent a fraction of the total costs of the project concerned. They will also amount to but a fraction of the cost which would be involved in implementing a "market" evaluation. Instead of asking whether researchers have done what they were asked to do, it is more appropriate to try to introduce accounting procedures which will ensure that the directors of research units have created sections which hum, which attempt to focus policy makers' and the public's attention on important issues, and which provide some answers to some important questions (even if overlooking a number of other important questions).

If researchers, like other public servants, need to have considerable security of tenure, how is their work to be oriented, and how are high levels of enthusiasm, dedication and innovation to be released among them? Sack doesn't quite get round to discussing this issue—but the answer is: "by the same procedures that public servants—and the users of research—need to be held accountable". Directors of research need to be held accountable for releasing high levels of enthusiasm, initiative, concern with high standards, and concern to meet the needs of the clients of the research among their staff. Researchers need to be held accountable for the soundness of their work, for stimulating and contributing to debate about important issues, for developing new understandings, and for developing the crucially important tools which are required to run the educational system effectively.

A Constituency for the Work. As Sack argues—and as is also apparent from Middleton's commentary—it is important for personnel who work in policy monitoring, evaluation and development units to establish a constituency—of politicians, administrators, practitioners, and the public—who appreciate their work and who are prepared to press for its continuation if embarrassing results are produced or if there is a change of administration. Policy researchers' contribution to political debate makes them particularly vulnerable; the process of monitoring and evaluation—however good the researchers' intentions and fair their statements—will often times be threatening to at least some of those who are evaluated.
Comment will be treated as attack. While all may have different reasons for doing so, conspiracy against evaluation and monitoring is easier than combining together to create a case for it and a climate in which the results can be used. From the policy researcher's point of view, it is not only criticism of current paymasters which is dangerous. So, too, is support for them - for these administrators may well be replaced by a rival - and perhaps critical - group in the future. Even criticism of lower-ranking officials may have serious implications because in a society governed by particularistic norms, such personnel may have contacts in high places - and evaluators may therefore find themselves under an obligation to find in their favour.

How is such a constituency to be established?

Part of the answer is, as has already been noted, to establish a network of contacts who have appropriate expectations of their work. (The phrase "appropriate expectations" is introduced because the expectations of applied social research tend to be far too high and far too diverse). The network of contacts with users may involve individual researchers sitting on policy committees or working with teachers and parents.

Another part of the answer involves the creation of mechanisms whereby people with particular needs can initiate appropriate enquiries.

Another part involves the creation of networks whereby interested personnel can participate in debate about the implications of the results of research-based enquiries.

Another is the creation of a network of clients who are going to use the work to see that public servants do their job.

There may also be room for formal reviews of user satisfaction with the unit’s work.

Two implications of these suggestions need to be highlighted: the execution of such activities requires considerable time from researchers - time which has somehow to be funded by the unit’s sponsors. The other is that such active participation in the process of policy formulation and development demands time from the users of the research - time to explore ideas, to reflect, to innovate, to consult, to participate in the wider framework of decision-taking. Such time needs to be provided for in the budgets of the action (not the research) agencies. Yet the provision of time for such activities relates directly to the wider question of the structures and the procedures which are required to promote the effective implementation, improvement and management of public policy which has formed the primary focus of this paper.

Staff Development and the Role of Consultants. We have seen that an expatriot researcher like Schwarz can provide the Bank with some of the information needed to assess the value of a particular project, and to identify the steps needed to improve it, more quickly and more effectively than can the, often inexperienced, personnel of borrowing countries.
Unfortunately, this process fails to address the more basic problem of helping to develop the institutions, expertise, and expectations which the countries concerned require if they are to run their economies effectively and take responsibility for their own economic and social development.

Not only is it important for researchers from the borrowing countries to discover for themselves the strengths and weaknesses of a variety of approaches in evaluation, it is also important for them to develop the subtle abilities required to use these techniques to generate useful insights and adapt them to their own needs. Consultants are required to provide help and reassurance while they are developing these qualities. The task of such consultants must be to encourage those concerned to persist in the face of difficulties and to encourage them to use the data in ways not identified in textbooks on research methodology or in the project design.

While there is also a role for travel and training, it would seem from our present vantage point that what is needed is not post-graduate training on the traditional—particularly American—model. It is to struggle with researchers working in policy evaluation and development units to try to find a way of solving the trainers' problems. It is to become enmeshed in the work of the unit. Clearly, if this is to be effective, the unit, and the personnel with whom the "trainee" is to work, must have been carefully selected to mesh with his own needs. The role of the trainees is to contribute their own understandings to international debate about educational processes and outcomes (and their limitations) so that they can sharpen up their understanding, obtain relevant feedback, and come to think of themselves as people who are expected to make original contributions and invent their own techniques. There is nothing like participation in the shambles of an international conference to boost self-images and encourage those concerned to make relevant contacts with the international research community from a position of strength. Under these circumstances, their contribution is one at the forefront of knowledge to the evolution of international understanding of what everyone is about. Training does not consist in telling; it consists in participating in the evolution of appropriate procedures and understandings hand in hand with the best mentors available.

As Middleton and Schwarz found, relationships with consultants are often unsatisfactory. As both report, some members of the staffs of the research institutes of the countries concerned asserted that they had nothing to learn from external consultants. While the strength of this feeling was, no doubt, in part due to the fact that they had to pay so much for the services of an external consultant, it is not entirely clear that it is appropriate for Schwarz to have ended up agreeing with them. Even the British banker, Rothschild, realises that it is of critical importance for social researchers to undertake activities which few of them have identified as important or feasible, which are of greater magnitude and significance than those which they envisage for themselves, and in relation to which they can conceive of asking for technical help. Schwarz's conclusion is only sustainable if it is either true that the borrowers can get their minds round the task which needs to be done or if the object of the exercise is in fact merely to enable the Bank to get good information on the effectiveness
of a particular project. The argument is not sustainable if the primary objective is to develop the borrowing country's capacity to undertake program evaluation and development for itself. If the latter is the case there are complex, and poorly articulated, tasks to be undertaken and those concerned need someone to provide continual help and reassurance as they venture into unchartered territory, try to identify the mountains to be climbed, find ways of scaling them, chart the resources available, and establish the complex social arrangements required for their exploitation.

If we are to envisage applied social researchers performing the role we have outlined in this paper, what kind of consultants are required? Where are they to come from? What relationships should the Bank encourage borrowers to make with the consultants they hire? And to whom do the consultants owe allegiance - to the people of the borrowing country, to their local paymasters, to the Bank, or to mankind in general?

From our present vantage point it is clear that a range of different types of expert is required. The necessary expertise ranges from expertise on the implementation of social management systems through to methodological expertise. One of the results of employing too few consultants and researchers, and employing them on an insecure basis, is that they tend to disparage the expertise of colleagues who are authorities in other areas of activity. But our assertion that a range of different types of expert is required is not the most helpful way of stating the need, because, as we have seen, a consultant is not best viewed as an expert who will tell those he is working with what to do. A consultant is best viewed as a colleague who is struggling to understand similar problems. His value, well illustrated in Schwarz's study, is that his experience is more likely to lead him more quickly to focus on critical issues, and to invent ways of obtaining relevant data and insights. It is not to communicate "techniques" or "methodology". It is an overstatement to say that methodology is unimportant; that what one needs to do is to develop an understanding of the problem before one starts, and then invent the methods which will enable one to collect some relevant data - but that statement nevertheless contains more than a grain of truth. The consultant therefore needs to be closely involved in the project and to understand its functioning and difficulties. He needs to be someone from whom social researchers will benefit as a result of working together over an extended period of time. The consultant must therefore be employed as a partner in a co-operative enterprise from which he, as much as his clients, will develop new understandings and insights. His task is to share his thought processes, not just his knowledge, and to provide encouragement and reassurance as researchers discover just how unlike textbook prescription the messy business of policy-relevant social research actually is, and struggle to get something useful out of it. His task is to encourage those he works with to struggle to clarify what they are doing and the reasons for their problems. His task is not to do their jobs for them. His task is to seek to establish the institutional frameworks and expectations which are essential if they are to work effectively.
While technician-type consultants demonstrably have a role to play in this framework, theirs is clearly not the dominant role. The need is for a gentler and more enduring relationship, for people who have the time, the patience and the political acumen required to influence widespread attitudes to social research in the countries concerned. The main need is for contact with people who have devoted large sections of their lives to researching topics which are of relevance in the countries concerned and to exploring their policy implications. It may be thought that it would be impossible to prise such personnel from their desks to assist in the work of researchers in LDCs. That is not necessarily the case, for such personnel will rarely have been able to pursue their own interests single-mindedly. They will usually be experienced at harnessing the prevailing breeze and accomplishing something which both suits their purposes and satisfies those who are anxious to obtain short term advice and information. But they will not be research whizz kids who are willing to run before the wind and rush in almost any direction in response to a financial bait. Nor will they usually be academics who are anxious to be seen to be associated with the latest fashion in their quest to advance their careers.

The question of consultants' and researchers' contracts is therefore problematical. It is plain for all to see that, whilst it cannot be claimed that nothing has come of the contract-research movement in the West, the most valuable contributions to policy evaluation and practice have come from people who have been able to pursue their interests, often on a relatively small scale, over many years. Contract research tends to generate vast paper mountains, and be an expensive way of gaining the relatively low level insights which tend to be produced. The headlong rush into activity frequently ends up on the rocks. Projects with "clear objectives" tend to reach conclusions which are of limited value when conditions change. Research pitched at more fundamental and enduring levels tends to have more long-term significance and, appropriately organized, can provide the insights and tools which are needed by policy makers - from parents to Ministers of Education - to implement policy. People who have been involved in policy-relevant research over an extended period of time - and taken the time required to develop the policy implications of their work - are, therefore, preferred consultants. And they need to be recruited to work with the projects concerned over an extended period of time - in the expectation that they will both enhance their own understanding of issues of concern to them, share their emerging understanding with those they work with, assist in improving the host country's institutional framework, and develop the confidence and competence of the host country's personnel. In recruiting them, attention needs to be paid, not only to the arrangements made with the institutions of the host country, but also to the arrangements made with the institutions for whom the consultants normally work. The consultants who are needed will usually be established policy researchers who have work to continue and to report; they will have correspondence to maintain; they will have careers to pursue. The costs of enabling them to pursue all these activities form a legitimate part of the budget which the borrowing country can be expected to pay.
What has been said about consultants is predicated on the assumption that the major task of the consultant is to establish the institutional framework, range of expertise and experience, and patterns of expectations which are required if the monitoring and evaluation procedures which are crucial to the effective operation of a modern economy are to be developed and established. In this context, it might behove the Bank to encourage borrowers to recruit the aid of a number of people who have spent time thinking about the understandings, tools and procedures which are required to run a managed economy effectively. Likewise it might be useful for them to recruit some who have proved themselves to be particularly effective in managing policy-relevant research institutes which have been shown to be particularly cost-effective in terms of generating useful new insights and understandings in western society.

Conclusion

It is apparent from our discussion that the primary task of the Bank must be to help borrowing countries to develop the structures, processes and expertise which are required to administer a managed economy effectively. Its role cannot be limited to that of assessing the value of particular funded projects. While superb studies conducted by ex-patriots have a definite role to play in informing Bank decisions, and clearly have a role in the more general process of promoting the economic and social development of the countries concerned, such studies cannot form more than a fraction of the whole.

Nor can the Bank even content itself with attempting to establish evaluations of particular projects; to answer Middleton's question about whether the Bank should be involved in establishing and monitoring innovations - it must, indeed, embrace the role of promoting the necessary programs of innovation, complete with monitoring activity.

To this end there is a need to promote discussion of:

- the nature of modern economies and the understandings and procedures which are required to run them effectively;

- appropriate accountability procedures - accountability procedures based both on utilization of professionally-developed instruments of accounting and on frameworks of public surveillance which could be harnessed to give teeth to the results of professionally-conducted social accounting exercises;

- the role of social researchers and the links to be established between them and the users of research - defined to include the public, local officials, central administrators, and politicians.

When initiating a specific proposal for policy development, monitoring and evaluation, our discussion suggests that Bank staff might usefully bear in mind the wider issues discussed in this paper. They should, in particular, consider:
the institutional base from which the policy researchers are to work;

the career structures and security of the researchers and the continuity in their programs of research;

the links to be established between the researchers and the users of research - including in the latter, personnel at all levels from the general public to politicians;

the ways in which a constituency, which will press for continuity and extended coverage in the work, is to be established;

the differing needs of the various users of the research;

the feedback loop which will enable the insights gained to be reflected in alterations to the design of the intervention;

the feedback process to be used to ensure that the insights generated in the course of the research are used in the design of further programs;

the range of research styles and studies which are required (including some studies addressed to fundamental long-term issues such as finding ways of measuring the most important outcomes of the educational process) and providing legitimation for this range of studies;

the range of consultants with different types of experience which is required - and the briefs of each;

the role of ex-patriot consultants and the way in which their home institutions are to be interfaced with the units with which they will be working in the LDCs over an extended period of time;

the expectations which are built up in the borrowing country concerning the outcomes of the research and the role of consultants;

the expectations which are built up concerning the role of evaluation and evaluation studies in a modern economy;

the overall management structure of the project and the criteria against which the managers of the research are to be judged;

the overseas development of researchers (working with the institutions to which they are attached rather than being told how);
the variety of types of study needed; these will probably include: (a) some input-focused studies of the sort undertaken by Schwarz; (b) a variety of different types of outcome-focused studies, some short-term and some long-term; (c) some illuminative studies - studies of processes and the explanation of linkages with a view to discerning intangible outcomes; All of these should be initiated in the expectation that it will be necessary to draw lessons from dirty data rather than in the expectation that convincing proof (in the form of a few lines of figures) can be laid before administrators. Time to mull over, re-analyse, and reflect on the implications of the dirty data therefore needs to be built into the budget from the beginning.

The Bank should specifically take care to alert their borrowers to the fact that the U.S. contract research system is an extraordinarily expensive and inefficient procedure for generating the benefits it has yielded. A new framework of relationship needs to evolve and LDCs have the opportunity to experiment with alternative patterns. Attention should be drawn to the fact that the problem has centrally to do with the criteria against which policy-research units, their managers, and their staff are judged. It therefore has to do with the very problems of public accountability that the unit will itself be concerned with. The Bank should also take care to draw borrowers' attention to the fact that the costs of implementing an effective policy development and evaluation system are small compared with the costs of the systems they will be monitoring; further, that there is generally considerable inefficiency in public provision and that considerable savings can therefore be expected.

Finally, it is clear that duties ancillary to carrying out research - duties of an advisory, problem-identification, proselytizing, implementation and dissemination, and organizational (staff development; establishment of a constituency) nature - are very considerable. Indeed, they probably amount to the lion's share of the work to be undertaken. The implications for Bank funding will be obvious. Not so obvious is the desirability of identifying these activities as among those which the funded unit is to carry out.
References


