

34193 to Newman

Evaluation of the Pakistan Primary Education Project:

A Methodological Case Study

August 23, 1983

Paul A. Schwarz

TABLE OF CONTENTS

The Methodological Context	1
The Primary Education Project	5
The Research and Evaluation Component	9
The Evaluation Component Rethought	11
Conduct of the Reconnaissance	14
Problems in Providing Quality Education	15
Characteristics of the Inputs Being Made	18
Apparent Effects and Accomplishments	23
Important Process Elements	26
Important Contextual Factors	27
Implications for the Project in Pakistan	29
Implications for Other Projects	30
Record-Keeping Procedures	31
On-Site Reconnaissances	32
Periodic Progress Reviews	32
Special Studies	33

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?

	<u>Test of Potential</u>	<u>Test of Readiness</u>
<u>Necessary Inputs</u>	Type I	Type III
<u>Sufficient Inputs</u>	Type II	Type IV

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 1 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 1 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 school 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)...a 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 a 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 be 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 I.~~ 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 inutility 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 earliest 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 experiences of the project participants. The task was to extract it.

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 the 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 others?" 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 shuts 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 it 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 a 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 most 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 that 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 outcomes 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.

- (8) 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.

- (9) 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 sometime 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.

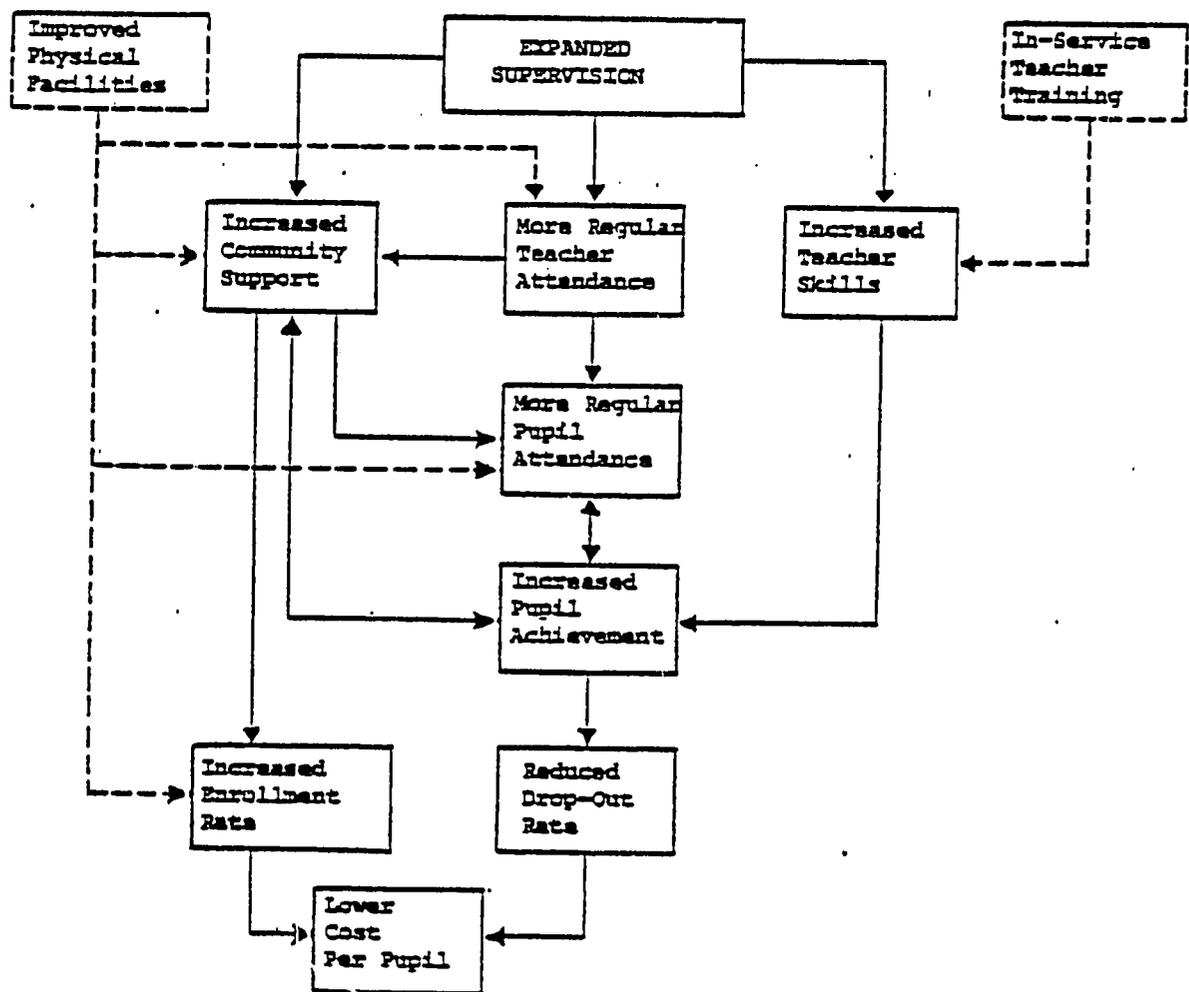


Figure 2. Schematic "rationale" of the supervision component

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 landlords, 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 accomplishments 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 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 to 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 teachers 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. 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.

B. J. H. W. B. J. H. W.

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 conclusions 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 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

New

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

New

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 phenomena 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.

New

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) would 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

VE

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.