

THE WORLD BANK

URBAN AND REGIONAL REPORT No. 82-~~20~~

11

EVALUATION OF PUBLIC SECTOR INVESTMENT PROGRAMS

DOUGLAS H. KEARE

May 1982

This report is being circulated to stimulate discussion and comment. The views reported here are those of the author, and they should not be interpreted as reflecting the views of the World Bank or its affiliated organizations.

INTRODUCTION

The activities of development lending institutions, such as the World Bank, imply tests by which their effectiveness can be gauged. Although the World Bank has always prided itself on its pragmatic, "learning by doing" approach to such matters, it has been surprisingly slow in systematizing the lessons of development experience so that they can be applied to planning. Evidence of this fact is the considerable gulf between estimated and actual costs and time periods required for project implementation. Further evidence is the number of effects of its interventions about which the institution cannot make confident judgments, even in the cases where these interventions have been going on for some years. In short, the World Bank lacks an adequate system for recording and evaluating these activities and their consequences to the societies within which they occur.

The need in this sphere was recognized in some measure years ago. The first recorded "stirrings" within the Bank date from 1967. Ostensibly, the evolution of new approaches to development then occurring kindled a new interest in evaluation. As these approaches were applied, the more complex projects they spawned confronted practitioners with the extent of their own uncertainty as to results.

With program development or expansion in important new areas, most notably rural and urban development, came the initiation of evaluation programs in 1972-73. As they have now been underway for more than five years, an initial stock taking is in order.

Those who think such an exercise has already been carried out should be quickly disabused. A staff Working Group was convened in 1978

to examine the status of monitoring and evaluation within the Bank and make recommendations. (Early discussions of the subject, and staff even now, tend misleadingly to treat monitoring and evaluation in concert. This practice will be avoided here, and monitoring will be discussed only when necessary for the sake of clarity.) The work of this committee -- which issued its report in February, 1979 -- was quite superficial. It failed to take a sufficiently hard look at either the weaknesses or the strengths of the experience to date. And although it could scarcely avoid the conclusion that not much was happening, it tended to impart the notion that all was proceeding more or less as it should.

This is hardly the case. Much confusion reigns within the Bank, and not alone there, as to what evaluation is or ought to be about, what it should aim to achieve, and how it should be designed and conducted. For present purposes, we may define evaluation in general terms as follows: evaluation comprises any set of analytical activities, embracing associated data generation and gathering, which aims to provide a record and appreciation of an institution's activities and their consequences.

It is the task of this book to clear away the confusion that persists on this subject: to spell out the pragmatic bounds of evaluation; to show that experience with data has given rise to an approach that is demonstrably viable and rentable; to assess the desirable program in light of institutional constraints; and to recommend guidelines for evaluation practice.

PART I

Chapter One:

DEVELOPMENT AND THE NEED FOR EVALUATION: THEORY AND HISTORY

Genesis of Evaluation

It is surely true that thought and practice in all fields, including development, are even in flux. Still, there are periods in which change is especially dramatic and focused. One such period in development practice began sometime in the late 1960s. Intellectual foreshadowing of changing practice were first apparent earlier in the same decade.

Prior to that time, development theory and practice had been firmly based on the principles of capital transfer, industrialization, and appropriate management of the public sector budget. Development strategy centered on direct investment in industry, and in industry-facilitating transport and public utilities. Markets were assumed to operate efficiently and without important distortions, and all were supposed to benefit through spread and "trickle down" effects. This was the world of Marshall Plan assistance. Its mode of development practice persisted well into the 1960s, and two-gap analysis (a la Chenery and Strout) was all that was needed to evaluate performance. ^{1/}

Already in the early 1960s, however, the premises of this rather simple theory of development were being questioned. More attention was being given to balanced agricultural-industrial growth, to

^{1/} Cf. Hollis B. Chenery and Alan M. Strout, "Foreign Assistance and Economic Development," American Economic Review, September, 1966, and Hollis B. Chenery, "Reply", American Economic Review, September 1968.

internal and external terms of trade, and, once the differences between effective and nominal tariffs were quantified, to the distorting effects of protection. Meanwhile skills shortages and related problems demanded more consideration of the potential for human capital formation, and people began to perceive the frightening implications of the population explosion. The first wave of change, motivated almost entirely by efficiency considerations, led eventually to expanded developmental lending for agriculture, education and population programs.

A second wave of change in development theory centered on a sometimes fundamental re-examination of industrialization strategies. The distinction was drawn between export led (good) and import substitution (bad) strategies. Many observers noted that markets did not work as efficiently as had been supposed. Even rapid development did not tend to solve employment problems, leading to questions about factor price distortions. "Trickle down" simply did not occur to the extent hoped for. These observations fed a desire on the part of some to do more about distributional issues, and convinced them that these issues had to be handled directly, rather than as secondary effects of development. One eventual result of these changes in thinking was the initiation of complex projects aimed at rural and urban development.

Changes in thought and practice tend to be continuous and gradual. However, in the case of development practice there was a watershed period, roughly from 1968 through 1972, coincident with Robert McNamara's first terms as president of the World Bank. During this period, the breadth and pace of change altered drastically. The changes

were presaged somewhat during the presidency of Mr. McNamara's predecessor, George Woods. ^{1/} It was he who suggested the grand assize which was to become the work of the Pearson Commission. Convened by Mr. McNamara in 1968, the Commission brought together in its report many questions and concerns that had been building throughout the 1960s and did much to establish the case for new directions in development practice.

The Pearson Commission was principally concerned with the amounts of development assistance (primarily capital assistance) and its distribution among channels and between countries. But it did pinpoint and clarify the related problems of rapid population expansion accompanied by slow employment growth, factor price distortions and the resulting excessive use of capital, the inequality of income distribution, and neglect of agriculture and the poor. In doing so, it served to coalesce the major arguments against the traditional approach to development problems, and to provide a basis for devising better approaches over the ensuing decade. ^{2/}

Four broad changes in World Bank lending to developing countries were set in train during the second half of the 1960s, and these trends have persisted up to the present. First was an overall increase in the magnitude of lending, from an average of a billion

^{1/} These changes involved increased emphasis on agricultural and educational lending.

^{2/} Commission on International Development, Partners in Development, Praeger, New York, September, 1969, pp. 58-61.

dollars in fiscal years 1965-69 to over ten billion in fiscal 1979, for combined IBRD and IDA assistance. Second has been the near doubling in the number of sectors covered by Bank loans, with the addition of energy, industrial development and finance, population and nutrition, tourism, urban development, and substantial non-project coverage to the traditional sectors of agriculture, education, industry, power, telecommunications, transportation and water supply. (At the same time, agricultural development has been extended to embrace rural development, and water supply lending has been made to cover sewerage.) Third, significant shifts in the relative magnitude of lending in the various sectors over time, have been led by the tripling of aid to agricultural projects, as a percentage of total bank lending. Finally, and perhaps most important, the changes in lending have established new directions in the Bank's efforts to increase growth and foster more equitable distribution in both poorer countries and the poorer strata throughout the developing world.

Annual Banklending under George Woods hovered around the billion dollar mark. With the appointment of Robert McNamara as president, the volume of lending swung sharply upward. In his first major address to the Bank (1968), McNamara announced plans to expand loans and credits over a five year period:

We in the Bank ... set out to survey the next five years, to formulate a 'development plan' for each developing nation, and to see what the Bank Group could invest if there were not shortage of funds, and the only limits on our activities was the capacity of our member countries to use our assistance effectively and to repay our loans on the terms on which they were lent.

As a result of this survey we have concluded that a very substantial increase is desirable and possible. 1/

In fact, the Bank raised more funds by borrowing in the first quarters of FY69 than in the whole of any previous calendar years, and went on to exceed McNamara's goal of doubling loans and credits over the following five years. 2/

Illustrative changes within sectors were those that occurred in the area of agriculture. For nearly twenty years, the Bank's strategy for agricultural development concentrated on large-scale irrigation works and other capital-intensive projects. In FY56, for example, some ninety percent of funds were committed to loans for heavy farm machinery and major irrigation canals. By the mid-1960s, the Bank's interest had shifted focus to the sector as a whole, with the aim of increasing output and reducing the heavy (balance of payments) burden of food imports. As early as FY65, a quarter of agricultural lending was devoted to multi-component projects. This reorientation continued through 1972. Although the preponderance of loans were still committed to irrigation and land development works as late as 1970, the proportion

1/ Annual Address to the Board of Governors, September 30, 1968, p.10.

2/ In terms of sectoral orientation, lending for agricultural projects rose from a total of \$405 million (or 11% of the total) during 1965-7 to \$1791 million (or 20% of the total) during 1971-2, and continued rising in both absolute and relative terms. Lending for education projects in the same periods rose from \$115 million (3%) to \$565 million (6%), and that for water supply rose from \$50 million (1.4%) to \$520 million (6%).

of projects bringing benefits to smallholders continued to rise significantly. Attention did not shift markedly to distributional and poverty alleviation issues until the 1971-75 period.

All these changes meant that projects became rapidly more complex, incorporating more components and taking on more objectives. As a result, the newer projects led the Bank rapidly into areas where its experience and the total available knowledge were both severely limited. The increasing size of the institution and scale of its activities forced recognition of the need for more systematic study of processes and consequences of development. The situation was prime for the introduction of methodical evaluation activities.

Evaluation Theory Circa 1972: Types and Purposes of Evaluation ^{1/}

The late 1960s and early 1970s mark a period of transition in the roles evaluation studies were to play in social programs in the United States. With the new programs comprising the War on Poverty in 1964, Congressional and public pressures mounted on policy makers and program administrators to provide data on the outcomes of their efforts to increase employment, improve education, deliver health services, and promote access to federal assistance for the poor. To some extent, academic researchers were already interested in the kinds of designs

^{1/} This date coincides with the Bank's initial investigations and planning for evaluation in the urban sector. It is also a good reference point for considering the designs of other evaluation programs established by the Bank. Although the program to accompany agriculture and rural development had its beginnings in 1968, the Bank neither settled upon an approach to the activity, nor committed significant resources to it until 1972.

being demanded of program managers. But political pressures resulted in more scholarly attention being paid directly to evaluation studies. ^{1/}

Goal attainment studies were a major thrust in evaluation efforts in the early 1970s. The aim of such studies was to determine whether and how efficiently a program met its objectives. Edward Suchman and Joseph Wholey, among other proponents, insisted that rigorous research design was a must for such studies, for without it the efficacy of program elements cannot be distinguished from the effects of extraneous factors. Suchman outlined a five-fold strategy for conducting evaluation, taking account of a program's (1) effort, (2) effectiveness, (3) impact, (4) efficiency, and (5) process. ^{2/}

Wholey and his associates at the Urban Institute offered another typology, devoting some attention to input-control systems, but more to output-measure systems. They specified three program-input mechanisms: monitoring, reporting systems, and cost analysis. Unfortunately, Wholey's treatment did not provide ways of linking these mechanisms to the specific output measures that he derived from program experiences in the Office of Equal Opportunity. He distinguished the following output measures: program impact evaluation, assessing the

^{1/} Peter Rossi, "Testing for Success and Failure in Social Action." in Evaluating Social Programs: Theory, Practice and Politics, P. Rossi and W. Williams, (eds.), New York, Seminar Press, 1972, p. 14, et passim.

^{2/} Edward Suchman, quoted in Joseph S. Wholey, et. al. Federal Evaluation Policy: Analyzing the Effects of Public Programs, The Urban Institute, Washington, D.C., 1970, p. 94. Wholey quotes from Suchman's path breaking volume Evaluative Research: Principles and Practice in Public Service and Social Action Programs.

extent to which a national program met its objectives; program strategy evaluation, rating the relative effectiveness of different components in a program; project evaluation, to assess the effectiveness of individual projects; and project rating, the comparison of various local projects in meeting stated objectives.

Two methods for program strategy evaluation were suggested by Wholey. One was to study the natural variation (or relative effectiveness) of projects in a national program. The other was to introduce "planned variations," similar in form to World Bank projects, on a smaller scale of implementation. Such variations would include field experiments, with careful specification of treatment and control groups, control over input and process variables, and careful measurement of input, process and output variables; and even smaller-scale experimental demonstrations without control groups, but with a design allowing for comparison of the relative effects of alternative treatments. ^{1/}

Suchman and Marvin C. Alkin went further, recommending that social programs and their evaluations be developed in a phased cycle leading from pilot, to model, to prototype, to institutionalized stages. These stages entailed first a flexible evaluation design, to assess the crucial factors leading to success or failure; then a controlled experimental design to judge effectiveness; and finally a continual feedback design for collecting and processing information in such a way that ultimately the "whole interacting system" could be

^{1/} Ibid., p. 24-27.

understood. ^{1/} Alkin stressed the usefulness of defining evaluation goals at the planning stage of a project, to enable evaluators to feed information to management coherently throughout the implementation process.

Process Evaluation: Michael Scriven offered a related, but finally different approach to evaluation studies. He coined a useful distinction between "formative" and "summative" evaluations: that is, those done to assist project development while substantive changes are still feasible, as opposed to those aimed at assessing a social program after it has been in operation. Scriven insisted that evaluation should not be confined the estimation of goal achievement, but should play an active role throughout project implementation, measuring input and output variables so that the processes by which goals are met can be understood. He saw evaluation as an activity informing the entire realization of a project, not just as the final phase of it. ^{2/}

Scriven's approach requires that researchers measure the degree to which a social program's objectives are met, but also that they investigate the unforeseen consequences of program activities. He, and other commentators, called for a "process" or "systems" model of evaluation to track the processes by which program organizations interact with their beneficiaries in the social environment. They

^{1/} Edward Suchman, "Action for What? A Critique of Evaluative Research," in Carol A. Weiss (ed.), Evaluating Action Programs; Readings in Social Action and Education, Boston, Allyn and Bacon, 1972 (1969), p. 56-64.

^{2/} Michael Scriven, "The Methodology of Evaluation," in Perspectives of Curriculum Evaluation, AERA Monograph Series, 1967, pp. 39-83.

argued that by viewing these interactions as a system evaluators would be better able to measure relative efficiency in different sectors and to determine how well organizations perceive and overcome obstacles to the delivery of services. This approach is one that seeks to identify causes of program problems as well as their effects.

Both the goal-attainment and process models for evaluation were developed in response to particular experiences with social programs. The most common theme arising from reports of these experiences is that program objectives were seldom accurately or adequately specified. Social programs are complex and often have multiple short- and long-term goals that are not easily scaled by direct or proximate measures. Lacking clear goal statements, evaluators had to negotiate with program staff to arrive at acceptable statements of objectives, or had to develop their own. In other cases, where program aims had been specified, the designs and indicators for gauging their attainment were open to widely divergent interpretations.

Here again the political dimension of evaluation studies comes clearly into focus. Program managers may have various predispositions towards an evaluation exercise. They may cooperate with evaluation efforts, offer little or no assistance, or actively campaign against such efforts. Whatever the case, managers will feel free to dispute evaluation findings, claiming privileged authority on questions of program objectives.

Formative evaluation is sometimes constrained by organizational aspects of a program, as well as by theoretical or

logical frameworks. Even in cases where administrators are initially willing to aid evaluators, the latter may have data requirements that interfere with regular program operations, need funds that program staff might prefer to invest otherwise, or come to interim conclusions suggesting modified operations with which staff do not agree.

Summative evaluations, on the other hand, are inevitably accused of culminating (or beginning to produce findings) too late to help improve program operations. Finally, there is no assurance that even meticulously designed and executed evaluations will be translated into new policy or management decisions. Evaluation is of marginal use if its findings fail to generate recommendations that can be put into practice.

Evaluation Methods

The record of evaluation results up to 1972 is not very impressive. Peter Rossi averred that evaluation research up to that time had been characterized by "vague goals, strong promises and weak effects." ^{1/}

The promise of evaluation research, from the practitioners' perspective, rested on the applicability or modifiability of controlled experimental designs to the demands of wide-scale social programs or development projects. These designs, which require randomized assignment of subjects to "treatment" and "non-treatment" or "control" groups, offer the soundest methodological basis for establishing (vectorial) measures of a program's effects on a treatment group. Proponents do not

^{1/} Rossi and Williams, op. cit., p. 16.

claim that such designs are definitive. They do claim that controlled experimental designs deter most significant internal and external threats to the validity of research findings. Rossi noted that:

two main positions [on the controlled design issue] can be discerned. There are those who center their attention on what are the obstacles to the use of controlled experiments in impact effectiveness studies; and in contrast, others who propose to abandon the controlled experiment model as either impractical or irrelevant, and who are trying to build alternative models. 1/

Critics of the model emphasize that randomization is difficult to arrange in real-world social experiments. They point out that experimental designs do not answer the need for program improvement because they tend to provide useful information only after a program has run full cycle, rather than at planning and implementation stages.

As Wholey and his associates have pointed out, where program resources are limited and service clients are screened "in" and "out" by selection procedures, a controlled sample is not impossible to devise. Their viewpoint responds to the misconceived objection that to relegate a subject to a control group is to deny him benefits:

Use of a randomly selected control group simply introduces a set of 'non-treatment' subjects in a roughly uniform fashion throughout the whole group of treatment subjects. This implies the need to modify the cut-off points of the selection procedure ... to satisfy a random selection requirement. Since selection procedures in social programs tend to be fuzzy at the margins anyway -- and often along much of the whole worthy-of-selection line -- the use of control groups

1/ Ibid., p. 29.

selected on a random allocation basis would not seem seriously subject to denial-of-benefit charge. 1/

Social program operation would then potentially be guided by tracking the progress of experimental and control populations.

The actual record of experimental design evaluations being applied to programs pertinent to development aims is very sparse prior to 1972. Even under the assumption that the results of this method were superior, research reveals that the instances where the necessary conditions prevailed were rare and tended to have been rather simple situations, far from the order of difficulty involved in controlling for housing improvements in the cities of developing nations. Peter Rossi argued that, as of 1972, there were "almost no examples of evaluation studies of current programs which have followed ... models (from the controlled experiment tradition) with any appreciable degree of fidelity." 2/ In his view, the literature recorded instead a variety of quasi-experiments and other tests that attempt the most suitable match of non-participating populations with those participating.

Quasi-experiments resemble true experimental designs, but do not control for all major threats to internal and external validity. They commonly take the form of non-equivalent control groups designs, in which characteristics of comparison groups are measured before, after, and sometimes during program implementation. Typically, challenges to the comparability of the groups in question arise over the criteria on

1/ Wholey, et. al., op. cit., p. 91.

2/ Rossi, op. cit., p. 29.

which they are compared. Alternatively, the interaction of selection effects with other criteria, such as maturation or regression effects, are held to be damaging to results. Groups are matched on the basis of what are initially thought to be relevant criteria, and if possible randomized before being assigned to test groups, a technique that increases the statistical accuracy of the design.

The fact that non-equivalent control group designs are often only as good as the matching procedures used in comparisons (and depend too on the quality of the data base and complementary observational procedures) can be illustrated with the results of three cost-benefit analyses of the Office of Equal Opportunity's Job Corps Program from 1969. The OEO's former director, Robert Levine, noted that an external economist calculated the benefit-cost ratio of the program to be about 1.5:1. The Job Corps' own analysts estimated 4:1. And the General Accounting Office came up with a figure of 0.3:1. The discrepancies derived entirely from the use of three different groups for comparison with the Job Corps enrollees, although it is possible that greater divergences could have resulted if the analysts had also disagreed about the actual effects on program participants. The case also highlights the fact that quantification in itself provides no assurance that benefits can be determined unequivocally.

Time series designs are a more common form of quasi-experiment applicable to evaluation purposes. They involve a sequence of measurements of program variables before, during and after the implementation process. Such measures permit evaluation of a social program's effects

through the identification of clear patterns of change (or observation of their absence). Although useful to a degree, these designs cannot control well for the coincident occurrence of program effects with non-program effects in the operational period.

Multiple time series designs can redress this weakness somewhat. These designs have often been employed in education and transportation settings, where pre-test, program-test, and post-test scoring indicators are readily available. If one can find an appropriate comparison group and take "readings" on it at the same intervals, the comparison of first differences can constitute evidence of program effects.

Donald Campbell and his associates used this technique in evaluating a crackdown on highway speeding in Connecticut during the 1950s. Fatalities declined immediately after the program began. However, because highway fatalities registered an up-and-down pattern in subsequent years, comparisons were made with four neighboring states not implementing a similar enforcement pattern. Because these states did not experience an equivalent drop in the number of fatal accidents, the evaluation appeared to corroborate program proponents' claims that it had produced the intended effect.

Some evaluations using nonexperimental designs have also been attempted. The formats have included: single program, before-and-after tests; after-only tests; and after-only with comparison tests. Such approaches are frequently apt for "formative" purposes in monitoring program progress, providing spot checks on adequacy of

service delivery, and preliminary assessments of program effectiveness. As "summative" exercises, they have many drawbacks because they cannot differentiate between factors internal and external to the program. Many federal government studies have relied on these designs owing to time pressures, emphasis on quick feedback, and the argument that it is better to assess each program on its own terms, since each is unique, rather than attempt spurious comparisons.

Systems and process models derived either from evaluators' long-term experience in organizational settings or from their outright dissatisfaction with the controlled experiment model. Robert Weiss and Martin Rein argued against the controlled model. They contended that in broad social intervention programs "there are so many different ways in which changes could be related to the program that a very great number of indicators must be included. Even then there is not assurance that some factor has not been omitted." They added the following objections: (1) "The situation is essentially uncontrolled. Communities are open to all sorts of idiosyncratic experiences, and comparison samples are usually not comparative;" (2) "The treatment is not standardized. Forms taken by broad ... programs differ in different communities in response to different needs and tolerances ... Experimental evaluation usually neglects the study of the intervention process itself ...;" and (3) "Experimental designs discourage unanticipated

information ... The question is not "does it work?" but "when such a program is introduced, what happens?" ^{1/}

Weiss and Rein's complaints derived from their experiences with an urban youth service program. The evaluators, who attempted to implement a controlled experiment design, received very little support from the program staff, especially after it appeared that the program's effects might be negligible. The evaluation design in this case attempted to look at individuals' responses to the programs, whereas the program itself was committed to institutional and agency change. Because of these conflicting interests, there was little positive interaction between the two teams.

Following their experience with an evaluation design that did not accord well with program circumstances, Weiss and Rein favored instead what they called "process-oriented qualitative research." The method has three features. The first is sensitive monitoring of a program's implementation by field staff who collect data on both institutional and individual changes. (Since programs tend to proceed by trial and error anyway, the evaluators can record successes and failures of various approaches.) The second feature is the observation of the beneficiary population "as a system, composed of interacting parts, which reacts organically to programs addressed to it." Finally,

^{1/} Robert Weiss and Martin Rein, "The Evaluation of Broad-Aim Programs: A Cautionary Case and a Moral," quoted in Marcia Guttentag, "Evaluation of Social Intervention Programs," Annals of the New York Academy of Sciences, 1973, no. 218: pp.3-4.

the design provides information in time for administrators to adjust program conduct for the best possible outcomes.

While the flexibility of this method is a strong point, Weiss and Rein provide few details on the actual collection, recording, and measuring systems to be employed. ^{1/} It is not clear from their description how the research differs from an effective project reporting and recording approach. That is, it is not clear how such research acquires evaluative content.

In her handbook, Evaluation Research: Methods of Assessing Program Effectiveness, Carol Weiss devoted some attention to a path model for evaluation analysis, a method that relies on diagramming the multiple-track sequence of events through which a program is to progress and its impacts are to be measured. Problems are anticipated by setting up checkpoints for monitoring how well program expectations are being met. Weiss notes that if program strategy shifts, the evaluator can follow suit and redesign measures for testing the delivery of services. Early construction of this kind of model can aid program personnel in testing how realistic their assumptions about operations are. ^{2/}

Edward Glaser and Thomas Backer defined their "clinical approach to evaluation" as one that assumes a research design must

^{1/} In Rossi and Williams, op. cit., p. 34.

^{2/} Carol H. Weiss, Evaluation Research; Methods of Assessing Program Effectiveness, 1972, pp. 50-53.

address a program as a whole with complex internal interrelations. ^{1/} They stress methods of subjective measurement, consultation, feedback, debriefing of program staff, and participation observation. The clinical method is intended to supplement others, depending on their particular merits in the situation to be evaluated, on the purposes of the evaluation, and on the problems of data collection.

The problem with all of these non-experimental methods is their failure to spell out the details of evaluation practice. They recognize problem identification as a matter of critical importance, but they offer general advice, rather than specific guidelines or procedures for tracking inputs and outputs and measuring impacts. So much attention is given to causes of program performance in the literature, that the issue of accurate measurement of program effects is insufficiently considered. In short, it is difficult to develop confidence by reading the literature that the methods will be operationally useful in specific situations or provide means of generalizing findings.

As of 1972, practitioners favored four different designs in their research. These ranged from strictly controlled experiments with factor-specific designs; to modified quasi-experiments needing some subjective adjustment of uncontrolled factors; to correlational or covariance techniques with some statistical controls in ex post factor measurements (the norm so far in studies of broad American social

^{1/} Edward W. Glaser and Thomas E. Backer, "A Clinical Approach to Evaluation," Evaluation I(1): pp.54-59.

programs); and the use of qualitative approaches, such as systems or process models, calling for analysis of interventions as a flow of negotiated interactions among policy makers, program staff, evaluators, and target populations.

What is surprising is how little real progress and dialogue there had been up to this point, despite much having been written on evaluation questions in a short period. Proponents of various approaches tended to refine their own positions to a tiresome degree, rather than to look for the stronger points in all approaches, in hope of finding suitable blends. There was as yet little hard evidence to go on anyway. Advocates of any kind of evaluation could point to very few studies actually documenting impact assessments even at the social experimental level. Many of the most important studies had been organized quite late in the game, such that by default, if not by design, they fell into the least rigorous third category above. Although researchers recognized the problems involved in specifying measures of program effects and justifying variables chosen as indicators, few had so far attempted to generalize their findings or to account theoretically for the chains of inputs, outputs, effects, and impacts. While investigators signalled the need to connect ex-ante, in-progress, and ex-post evaluation for improving program efficiency and efficacy, few operational means had been devised for this purpose. It could not have been otherwise, given that so few program administrators had seen fit to commission studies or assign personnel to address the questions evaluation can answer.

The World Bank: First Steps

The first discussions of evaluation within the World Bank, as reflected in exchanges of memoranda, date from 1967. These interchanges intensified in the following year. They were a part of the general surge in research and analytical activities in the Bank pursuant to Mr. McNamara's election as president, and his appointment of Hollis B. Chenery as what was then called his economic adviser.

At about this time, members of the Board of Executive Directors began to voice interest in monitoring and evaluation of Bank development activities. They were primarily interested in goal-attainment evaluation that would tell them whether Bank-supported projects were turning out more or less as expected. They were most concerned with possible research in the new sectors of lending and the so-called "new-style" projects, areas of scant experience about which many prospective participant governments were apprehensive. Bank staff, especially in what was then the Sector and Projects Studies Division, were meanwhile interested in improving appraisal procedures and measurements. ^{1/} To this end, staff concentrated on traditional lending sectors, particularly on the largest or most rapidly growing ones, where the applicability of their findings would be greatest. The recorded tendency of people, units, and institutions to call various activities by the same name -- evaluation -- has continued down to the present.

^{1/} It is difficult to pinpoint the origins of evaluation within an institution pervaded by the appraisal process and cost-benefit analysis, which are closely related to evaluation as it is defined in this text.

Unquestionably, there is less confusion on the matter now than at the outset. Even then, however, everyone's definition appeared to have one thing in common: a goal attainment orientation. The Bank's own institutional nature seemed almost to predetermine this fact. The project approach to development packages the work of the institution in discrete units that can be evaluated. Cost-benefit analysis constitutes a ready vehicle for the assimilation of evaluation findings. And the Bank's investment in appraisal made it worth diverting resources to evaluation to confirm or improve upon appraisal procedures. The goal-attainment bias also dominated the initial design of the Urban Evaluation Program in 1972, and this was to have some unfortunate consequences.

The economists in the Sector and Projects Studies Division started analyzing projects that had been approved several years earlier. The first study of this kind--"An Economic Reappraisal of a Road Project: The First Iranian Road Loan of 1959" (EC-147)--focused on highways and was completed September 26, 1966.

There were other early studies in this sector, including an "end-use" study of an Ethiopian highway project, and the "reappraisal" of an El Salvador feeder roads project that attempted to assess the impacts of improved transportation facilities on agricultural development. Thereafter, the major focus was on irrigation projects, vehicles for most of the Bank's lending in the agricultural sector. These investigations began with studies of two projects in Mexico and led in early 1969 to an internal working paper, "-Performance Evaluation

of Eight Ongoing Irrigation Projects," surveying activities in India, Mexico, Taiwan, and Turkey.

Already in 1968, the staff responsible for these early studies felt that there could be no satisfactory evaluation of actual project benefits without an appropriate system for designing, collecting, and reporting such data. In the absence of such a system the only alternative appeared to be in-depth case studies of a very few selected projects, utilizing inferential data and extended field investigations. It was hoped that on the basis of several case studies it might be possible to propose the general factors determining the success or failure of a particular type of project. It remains surprising that a companion program for generating types of data needed for better evaluations was not launched, or even apparently considered, at the same time.

The responsible projects departments were encouraging borrowers to build so called monitoring and evaluation systems into their projects. However, these early efforts were very much conditioned by Bank priorities, and as time was of the essence, the use of simple monitoring systems established institutional practice.

Monitoring and evaluation (M & E) components were written into approximately five projects annually during FYs68-70 and ten in FY71-72. The amount rose sharply as time went on. However, the Bank did not commit significant resources to the design of these evaluation processes, or event to data collection. Because it also assigned no

staff to supervise evaluation research, very little progress was made in evaluation practice up to 1972.

The research activities of the Sector and Projects Studies Division had by then included several reappraisals of projects, sector planning models, refinement of social cost-benefit accounting, development of evaluation methods for new types of projects, a highway design study, a port study, and an ex-post analysis of electricity demand forecasts. This background made it possible, following a major reorganization of the institution, to propose conducting evaluation activities in departments with potentially conflicting administrative interests.

Those who had spent time trying to learn from the experience of past Bank-supported development efforts believed they knew what needed to be done. The first step was to organize systematic reporting of data for all new projects, so that actual and predicted benefits and costs could be compared. The second, already being carried out, was to make intensive reappraisal studies of selected projects. This was not the only way to have launched an evaluation program, and it was hardly the best. This book details the process by which evaluation activities advanced and became integrated with World Bank's development efforts in the urban sector.

Chapter Two:

THE BANK'S EXPERIENCE WITH RURAL DEVELOPMENT PROJECTS-
THE COST OF INDECISIVE EVALUATION PRACTICE

In the late 1960s and early 1970s, the Bank was expanding its activities within the agricultural sector to include new areas of development. The new projects, such as the Lilongwe Land Development Program (LLDP) in Malawi, centered on expanding farm production by small holders and were considered to involve higher risks than previous development activities. The Bank had had little experience in designing appropriate inputs for small holders or mechanisms for delivering them and measuring their effects. Monitoring and evaluation components were to help fill gaps in the Bank's knowledge of the conduct of such projects.

Bank staff seem to have envisioned monitoring systems as systematic data gathering procedures for determining the progress of project performance beyond the capacities of typical financial accounting methods. Such monitoring systems were not limited in use to the new smaller-scale livestock and irrigation projects in the early 1970s, including projects in Mexico, Uruguay, Uganda, and India. In these cases, as in the Lilongwe project, monitoring systems were expected to yield valuable data from the farm level to assist in the formulation of improved operating procedures and even of agricultural policy. A review of the relevant documents discloses inconsistencies in nomenclature. Both "monitoring" and "monitoring and evaluation" units were concerned with in-course data collection and analysis, and potential project goal re-orientation. Repeated reference is made in

this chapter to "monitoring and evaluation" both because it is the term used most often in the documents, and because the chapter recounts in part the process by which "evaluation" has come to cover the activities formerly grouped under "monitoring" and "monitoring and evaluation".

Prior to 1973, practitioners in the agriculture and rural development sector of the Bank had sought to improve project performance in a variety of ways. Project monitoring and evaluation components represent only one of a range of economic research activities. The Sector Working Paper of June, 1972 notes that sector survey missions were carried out in all of the new development regions to "identify and prepare promising investment opportunities." These initial explorations apparently took precedence over assessing and improving actual project implementation designs. Where pre-project survey missions did not absorb staff resources completely, those available were employed in post-project re-appraisals of project activities. Whilst the Working Paper mentions that Bank research was conducted both to "support operations" in the sector and to "establish a factual or conceptual basis for policy," the investigations concentrated little effort on in-course tracking and correction of project inputs and outputs.

As early as FY67, the Bank energetically supported efforts to set up a project capacity for measuring cropping patterns and yield changes in the rainfed area development program. It pressed for stronger investigation on forecasting production targets in the preparation stage than had been proposed, and for additional information from a UK-funded pilot project initiated near Lilongwe in 1965. These

concerns in fact contributed to a fairly strong data baseline from which an evaluation unit might work. Yet, once having provided this design support, the Bank's "interest in the evaluation process seems to have dwindled during implementation," according to the LLDP Project Performance Audit Report (1975). The Operations Evaluation Department (OED) commented in the same report that "the Bank did not pay enough attention to delays in setting up the [Evaluation] Unit, nor to the solution of staffing problems, and made no effort to see that conflicting objectives were resolved and that data of high quality were analyzed."

The work program undertaken by the LLDP Evaluation Unit was highly ambitious. It sought information in five categories: (1) baseline data from "developed" and "underdeveloped" program areas for verifying project assumptions; (2) survey data for in-course program monitoring; (3) survey data on achievements after completion of the program's intensive development phases; (4) specific data on subsectoral components, such as credit, for reformulating operational guidelines after first phase completion; and (5) specific data on enterprises such as livestock-raising, for which little data had existed in program planning stages. Yet, due to bottlenecks, actual reporting was limited to occasional papers on project activities and the partial collection and analysis of data on land classification, cropping patterns, farm production and consumption, and some subsectoral operations such as credit and extension. Although the Unit won praise for attempting to pin down its terms of reference to manageable studies, it was not able

to gather crucial farm management data in such a way as to make it useable in the short run to Project Management. It was also unable to establish that the highly detailed socio-economic information collected as part of an intensive small-scale survey would be representative of comparative data over the long (13 year) phasing of project implementation. Due to other project constraints these initial evaluation studies were abandoned.

Technical debate concerning the validity of the data flared too within and outside the Bank. Baseline data were charged with being too high, failing to take into account weather and pest effects. OED reports claimed to have found "disappointing yields" in their own survey of average yield estimates, despite conflicting observations of visibly improved husbandry. Finally, sampling procedures were thought to be inaccurate, the survey having been insufficiently broad.

In an Employment and Rural Development Division follow-up study, Dennis Anderson made significant efforts to answer these charges and to solve the "yield puzzle." He showed that the baseline estimates were not overly high, the difference between the contending rates apparently falling within the range of expected production increases. He went on to demonstrate that OED's average yield measures camouflaged differences between "developed" and "underdeveloped" areas, which reinforced the belief that yields had increased in response to improved farming techniques. Anderson also showed that the sample size, rather

than being increased to ensure internal validity, could have been halved and still produced reasonably reliable results. 1/

This debate was evidence that, as of 1975, the Bank (and other development institutions) still had not defined a set of appropriate roles and tasks for monitoring and evaluation units in project implementation, nor provided contingency plans for M & E activities in the event of staffing problems or other delays. In an overview of rural development in Africa, Uma Lele pointed out that the Lilongwe Evaluation Unit had stopped short of drawing connections between key program variables, such as credit provision, farm income, and default rates. 2/ Her discussion was critical of the Bank's failure to consider systematically the ways in which information on component and overall project progress were to be handled by management. The institution had not sufficiently specified the indicators to be used in monitoring input deliveries, or the steps to be taken to assess the causes of bottlenecks. Could M & E units help projects reach objectives best by restricting themselves to baseline and updated data collection and analysis, or by corroborating data produced by management information systems through independent checking mechanisms? Should they concentrate on studies of project subgroups to see if appraisal assumptions proved to be correct? And if a combination of these approaches were most advisable, what factors justified giving one priority over another? It was not until the late

1/ Dennis Anderson, "Fluctuations of Maize and Groundnut Yields in the Lilongwe Land Development Program," pp. 1, 17-18, 20-22.

2/ Uma Lele, The Design of Rural Development: Lessons from Africa, pp. 95ff.

1970s that experiences in other projects, such as the PIDER program in Mexico and the multi-state Agricultural Development Program in Nigeria, convinced Bank staff at all levels that evaluation activities in its agricultural and rural development sector needed reorientation and redefinition.

The Third Mexican Livestock and Agricultural Project:
Consequences of Monitoring without Preparation

The Bank's handling of the Lilongwe Program Evaluation Unit may have been indifferent after its establishment in 1969, but its policy seems to have been unrealistically optimistic in proposing a farmer-based monitoring system in the Mexican agricultural sector in the early 1970s. Because the third Mexican livestock and agriculture credit project was very large in scale, the Bank proposed that FIRA, the agricultural trust implementing the project, institute a monitoring system to determine whether the project was achieving its aims and how it might be altered to become more effective. FIRA, an institution combining capacities in credit analysis and technical assistance, was expected to advise both participating (loan administering) banks and farmers on the implications of the monitoring system's findings. As early as 1970, the Bank's appraisal team on the project devised an extensive series of complex farm models to yield physical and financial projections. These models were created not only to capture on-farm developments for FIRA analysis. Replications of the models were to be distributed to the large-scale farmers themselves, and explained by FIRA staff, to "assist the producer in monitoring the impacts of his investments, and seeing beforehand the financial implications of

achieving better or worse results than those predicted, particularly as determined by changes in management practices." 1/

It is not clear in retrospect whether the responsible Bank officers took time to consider systematically how the proposed monitoring system was likely to work out in practice, as the proposal used the very term "monitoring" in an idiosyncratic way. OED's detailed report on the conduct of the Third Loan, for example, is unable to explain how the Bank's models were to apply to the interpretation of FIRA's surveys of loan fund use on large scale farms. This report also fails to specify how FIRA staff were to convert the data into "on-the-spot" operational or policy recommendations for individual farmers. Apparently, the Bank's expectation was that FIRA would use the 25 crop-and-livestock models to collect information from a sample of participating farmers, analyze findings for initial aggregate effects on the farm sector, and finally disaggregate them into recommendations on farm management for the land owners.

In practice, however, FIRA did not follow the course evidently envisioned by the Bank. In OED's view, while the models "were useful as illustrative models for project evaluation to FIRA, which was training its technicians in these techniques, it remains less clear whether the models were genuinely useful either for estimating the rates of return which different investments were expected to achieve or for identifying and influencing the types of investments made." 2/ The actual survey

1/ Project Performance Audit Report, April 27, 1977, p.57.

2/ Ibid, p. iii.

carried out by FIRA was pretty much a pilot effort to evaluate immediate impacts of the loans made to farmers. The results (devised from a non-significant but "illustrative" sample of 80 firm-type landholdings) indicated that desired increases in net farm assets had been achieved through intensive rather than extensive use of pastures, but could not account for how these changes had occurred. Nevertheless, FIRA did become more interested in expanding the so-called monitoring system to include a larger sample. FIRA did so, as the Bank noted, principally because the increases would boost its prestige in Mexico. In the Bank's view, the survey was of extremely limited value until FIRA could demonstrate an understanding of why the changes had happened as they did. In its OED report, the Bank also concluded that FIRA had not concerned itself enough with providing for farmers the sort of technical assistance that would encourage them to keep farm records, measure changes, and thus understand why particular recommendations would benefit them as well as the agricultural sector as a whole.

Yet the Bank itself had been reluctant to provide the sort of intensive preparation FIRA would have needed to create the farmers' monitoring or "self-evaluation" system. Although the OED's third PPAR report advised that "it might be wise for the Bank to establish a group which could more systematically work with the lending agencies to implement the monitoring systems desired," ^{1/} the fourth report indicates that FIRA continued to experience difficulties in convincing

^{1/} OED, Third Mexican Livestock and Agricultural Development Project Performance Audit Report, April 27, 1977, p.47.

farmers of the value of record-keeping and in providing technical assistance to them. FIRA has still not produced an account of the causes behind the continued progress in expanding net farm assets and appropriated land use. Because of the costliness of such evaluative analysis, "FIRA believed that the results, when obtained, would be valuable principally to the Bank, and thus that the Bank should pay." ^{1/} OED concurs in part, saying that the Bank has given insufficient attention to the final costs of the information it deems desirable, on the part of the borrowers. In OED's view, the Bank has also failed to justify why it is pressing so hard for the noticeably loose formulation of the actual "monitoring and evaluation" system in the fourth loan. That term is not even used until the fourth report, and reflects the lack of consideration given to the design and purposes of the impact-determining effort.

The Formulation of Monitoring and Evaluation Procedures

As early as 1972, a Rural Development Unit was proposed to clarify trends and assist in operating integrated smallholder rural development projects. But the principal impetus for creating such a unit came in the aftermath of Mr. McNamara's Nairobi Address of September, 1973. In July, 1974, Warren Baum explicitly responded to the evaluation theme of the Nairobi speech by commissioning "A Control/Monitoring System of Rural Development Work." The Rural Development Sector Policy Paper of February, 1975, formally committed

^{1/} OED, Fourth Mexican Livestock and Agricultural Development Project Performance Audit Report, July 3, 1979, p. 13.

the Bank to monitoring and evaluating in-course project work in the following declaration of intent, which admitted that the Bank's expertise in this area was limited:

Because the Bank's knowledge and experience of how best to help the rural poor raise their productivity and improve the quality of their lives [is] limited, it is necessary to:

1. Build a degree of flexibility into projects so that modifications can be made as experience is gained.
2. Devise evaluation systems in order to (a) control and monitor the extent of deviations from expectations, and (b) learn the lessons of experience. But such systems are expensive and governments are naturally reluctant to tie up scarce human and financial resources in what might be regarded as sophisticated and esoteric monitoring systems. Such systems are necessary, not because aid agencies want them, but because they ought to be an integral part of the internal management control structure. If they are introduced for this purpose, they can facilitate supervision by governments and assistance agencies, and help in learning the lessons of experience.

It is unclear whether the call for institutionalizing M & E systems arose from the first few evaluation efforts or from the lack of post hoc analyses recording how projects had actually run. Dennis Anderson's progress report on issues in the monitoring and evaluation of rural development projects looks at aggregate needs for improving in-course reporting without covering qualitative concerns of individual projects in any detail. Until 1975, the Bank had tended to take stock of its project work largely through ex-post evaluations, either in the form of OED's Project Completion and Project Performance Appraisal Reports, or through surveys of completed projects, such as John de Wilde's Experiences with Agricultural Development in Tropical Africa (1968) or Uma Lele's The Design of Rural Development (1975). Valuable as these exercises are, they concentrate primarily on the

matching of ex-ante and ex-post in project predictions and results, or on cross project comparisons of specific production or institution-building activities. Neither effort reveals the ways information received and processed by project management led to the decisions that actually determined each project's course.

January, 1976, saw the formal establishment of a Rural Operations Review and Support Unit (RORSU) to address this problem. RORSU was assigned responsibility for assisting project managers and Bank operational staff in the design and implementation of M & E systems for agriculture and rural development projects. The Unit first devoted its efforts to reviewing the experiences of M & E staff in Malawi, Brazil, Northern Nigeria, and in the nationwide Mexican program PIDER, highlighting the common problems of overambitious terms of reference, and lack of continuous staffing and data analysis capability. RORSU provided direct field support to 17 projects in its first two calendar years. In addition to holding seminars for Bank personnel, RORSU also co-convened an international conference on M & E systems in Copenhagen, where much stress was laid on the need to establish a "hierarchy of objectives" among project inputs, outputs, effects, and impacts and the integration of M & E design with management information systems as a whole.

In RORSU's conceptual framework, the information systems needed to coordinate the tracking of the project process are comprised of four elements: (1) financial accounting and auditing systems, (2) a management reporting system, (3) monitoring, and (4) evaluation. The

management reporting system is the lifeline of a project, in RORSU's scheme, the continuous flow of information on the financial and physical progress of project activities, representing the basic data sets managers need for implementation. Those data are used in turn by monitoring systems, and supplemented with other data, for the timely gathering of information on inputs, outputs, and complementary activities that are critical to the attainment of project objectives. The management reporting system provides a record of ongoing events, and monitoring provides the interim system of squaring those events with expected or planned ones.

Monitoring is linked with evaluation by its provision of part of the information necessary to assess and adjust "policies, objectives, institutional arrangements and resources affecting the project during implementation." ^{1/} While "monitoring requires a simple system that provides continuous feedback of key indicators on project progress...based on simple approaches for [information] collection," ^{2/} evaluation efforts may require generation of additional information over a longer time span than monitoring to fulfill its service to management. To accomplish their individual functions, monitoring and evaluation efforts require different but complementary sets of indicators to gauge degrees of success in the meeting of project objectives. Evaluation builds upon monitoring, answering questions that stem from information provided by the management reporting system, by monitoring, and by its own efforts.

^{1/} RORSU, "Monitoring and Evaluation of Rural Development Projects: An Early Assessment of World Bank Experiences," March, 1978, p.4.

^{2/} Ibid, p.4.

RORSU's scheme gears monitoring and evaluation systems to verify the attainment of measurable project objectives, defined as follows: Project inputs are the resource quantities necessary for a project, such as capital, manpower, and technology required to irrigate a certain number of hectares of land. Project outputs are the physical outcome of the project inputs, such as the number of hectares of land actually irrigated. Project effects are the outcome of the use made of project outputs, such as the agricultural yield per hectare of land irrigated. Project impacts denote the changes in living standards resulting from project effects, such as an increase in farmers' incomes.

RORSU's approach is a direct response to the degree of complexity common to most rura development projects. Of all its mandates, RORSU perhaps take most seriously its responsibility to devise or revise information collecting and analysis systems only to the degree necessary for them to be handled adequately by project or government personnel. It frequently quotes Robert Chambers on the need not to collect unnecessary information: "It requires experience and imagination to know what is not worth knowing, and self-discipline and courage to abstain from trying to find [it] out." ^{1/}

In an early formulation of its M & E conceptual framework, RORSU attempted to match the levels of objectives not only with verifiable indicators and M & E stages or "means of verification," but also with important assumptions underlying the programmed attainment of

^{1/} Ibid, p.5.

objectives. Here, M & E units would have had to chart the course leading from actual inputs back to pre-implementation suppositions concerning input-output linkages; from actual outputs back to presumed input-to-output and output-to-impact connections. It is not clear why this ambitious fitting of assumptions to objectives was abandoned in practice, though its omission may simply reflect the concern for tightening up the scope of M & E activities. It might also be that the formalizing of assumptions and the tying of them to project objectives was a matter thought to be best handled by the Bank itself, either in appraisal or in strictly ex-post studies.

Whatever the reasoning behind the more circumscribed framework, RORSU has concentrated its own extended proposals for M & E systems (1) to specifying levels of objectives in project activities and results; (2) to seeking indicators that can serve as objectively verifiable measures or proxies of those objectives; (3) to identifying users of project information for whom reporting procedures must be devised; (4) to recommending data collection and analysis systems fitted to local administrative and processing capacities; and (5) to commenting on optimal organization of M & E units relative to the rest of project structure. These procedures have been outlined for two different cases: monitoring and evaluation of nutritional interventions, with particular reference to alternative delivery systems in Brazil, and of agricultural extension projects, with special reference to India. The latter case will illustrate the strengths and weaknesses of RORSU's approach.

A Staff Working Paper outlines the functions of an M & E system for agricultural extension projects. ^{1/} The paper was occasioned by the expansion of extension services in India from the command area to the state level. The monitoring and evaluation system was devised to help the state agricultural authorities oversee the activities of thousands, rather than hundreds, of newly contracted extension workers. The project management reporting system, upon which M & E efforts are built, is based on the diaries of extension workers, including input records on the numbers of farmers contacted, output records on farmers adopting recommended practices, and so on, as well as project records on extension staffing, extension training programs, demonstration equipment, and the like. Here monitoring involves such indicators as the content, frequency, and quality of the visits of the extension workers. Evaluation deals with the subsequent performance of contacted farmers, as measured, for example, by yields, cropping intensity and patterns, spread of recommended practices from contact to other farmers, etc. (See Figures 2.1 and 2.2).

While careful measurement and costly outlays are required to determine yield totals, Cernea claims they are less difficult data to collect than income figures would be. He avers that farmers would not report income accurately, even to extension workers who resided for a time in the farmers' villages, and previous research experience suggests that he is right.

^{1/} Michael Cernea & Benjamin Tepping, "A System for Monitoring and Evaluating Agricultural Extension Projects," Staff Working Paper #272, December, 1977.

Figure 2.I

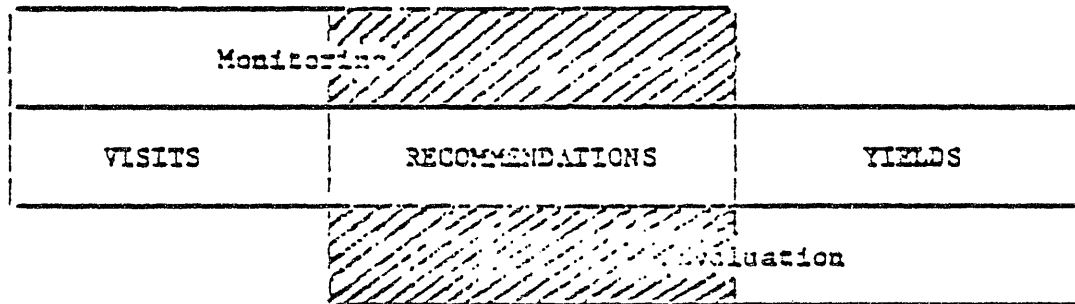
List of Indicators for Monitoring

<u>Objective: Institution Build-Up</u>	<u>Information Sources</u>
(1) Staffing of Extension Organization	- Reporting
(2) Selection of Contact Farmers	- Ad hoc study and reporting.
(3) Training (role learning)	- Reporting
(4) Physical Equipment	- Reporting and Accounting
<u>Objective: Extension Performance</u>	
(1) Degree of Exposure to Extension - Farmers reached directly - Farmers reached indirectly	Monitoring sample survey
(2) Quality of Visits	Monitoring sample survey
(3) Farmers' evaluation of T & V	Monitoring and harvest survey Village studies
(4) Adoption of farm practices	Monitoring sample survey Harvest study Village studies Farm Practices ad hoc study
(5) Role Behavior (VEWs, AEOs)	Ad hoc studies Monitoring survey
(6) Training (quality)	Ad hoc study Study on VEWs

List of Indicators for Evaluation

<u>Indicators</u>	<u>Data Sources</u>
(1) Yields of major crops	Harvest survey
(2) Cropping intensity and patterns (changes)	Harvest survey
(3) Area under HYVs	Reporting
(4) Spread of key practices	Monitoring and harvest surveys Village studies
(5) Amount of purchased inputs (fertilizers, pesticides)	Aggregate statistical information from distributors, etc.
(6) Credit use/recovery	Cooperative/Bank statistics

Figure 2.2



As can be noted from the above design,

- the VISITS will be the main concern of MONITORING
- the YIELDS will be the main concern of EVALUATION
- the RECOMMENDATIONS (adoption of) will be the concern of both MONITORING and EVALUATION.

Source for Annexes 2-4: Michael Carnea and Benjamin Tepping, "A System for Monitoring and Evaluating Agricultural Extension Projects," Staff Working Paper #272, December 1977, pps. 20, 22, 24, 101, 102, 107, 109, 110.

In Cernea's design, M & E efforts concentrate on two kinds of studies: sample surveys, and ad hoc or village studies. Monitoring surveys and spot-check studies are conducted principally during the growing season. They continue as long as data need to be collected on such topics as the adoption of recommended practices and interactions between agents and farmers, although timeliness of the information for project management also limits such studies. Evaluation surveys are conducted relatively rapidly during harvest time, supplemented by longer village studies investigating constraints and options in the spread of key practices, contact farmers' social networks, changes in cropping patterns, etc.

Cernea tries to keep the M & E surveys as brief as possible, limiting questionnaires to roughly two pages each. (See Figure 2.3 and 2.4.) Control groups are not considered feasible for these projects, due to external factors, such as the variability of weather among different villages, and to the project's objective of having the sampled farmers contact others to spread the innovations. Estimation of the diffusion effect is deemed more important in this case than changes in behavior between contact and non-participant farmers.

Finally, Cernea recommends that the M & E Unit be separate from the extension organization itself, preferably reporting directly to the state agricultural secretary, so that independence of operation can be assured. Ideally, the unit in each state would be tied to a Central Evaluation group in the Indian central government which would be responsible for quality control and technical advise to the units. The

Figure 2.3

DRAFT QUESTIONNAIRE - MONITORING SURVEY

Contact farmer's name _____

Farmers' Group identification _____

1. (a) What is the name of your VEW? _____ known
 unknown
 uncertain

(b) Which is the usual day of the week for the VEW's visit? _____

2. When did the VEW last visit your farm or some other farmer's fields with you present? _____

3. How many times in the last 4 weeks did the VEW visit you? _____

4. How many group meetings with the VEW did you attend in the last 4 weeks? _____

5. (a) How much land do you operate? _____ (Acres, Hectares, etc.)

(b) How much of it is irrigated? _____

6. Please tell me what practices does the VEW recommend for this crop season?

Description of Practice	Area Adopted	Area will Adopt	Extent of Adoption or Reason for Non-adoption

(Continue the list on the reverse side of page)

Figure 2.3 (Contd.)

7. Do you expect to have increased yields on areas on which you applied the recommended practices? ~~yes or no~~ If yes, how much? _____
8. How many farmers do you know who have learned the recommended practices from you? _____
9. Rate the usefulness to you of the agricultural extension program:
- (1) Extremely useful
 - (2) Quite Useful
 - (3) Useful
 - (4) Not of any real use
 - (5) The recommended practices are wrong
10. Please give any comments or suggestion the farmer would like to pass on.

Field Investigator

Date of Interview

Figure 2.4

DRAFT QUESTIONNAIRE FOR THE EVALUATION SURVEY

Crop _____ Farmers' Group Identification _____

District _____ Farmer's Name _____

Is the field on which the cropcutting was done irrigated? _____
yes no

Is the crop a high yielding variety? _____
yes no

Interview Section

- 1. Are you a contact farmer? ___ Yes ___ Not now, but earlier ___ Other
- 2(a) What is the name of the VEW for this area? _____
- (b) Which is the usual day of the week for the VEWs' visit _____
- 3. On how many times in the last four weeks have you talked to the VEW? _____
- 4. How many group meetings of the VEW have you attended in the last 4 weeks? _____
- 5(a) How much land do you operate? _____ (Acres, Hectares)
- (b) How much of it is irrigated? _____
- 6. Please tell me what were the most important practices the VEW recommended for this crop season?

Description of Practice	Area Adopted	Extent of Adoption or Reason for Non-adoption

Figure 2.4 (Contd.)

Which new practices did you use on the plot selected for the crop-cutting and how did you learn about the practice? (Mark all applicable columns.)

Practice used (Code from standard list)	S o u r c e			
	The VEW	A Farmer	Own Observation	Other (explain)

8. What would have been the difference between your present yield and what you would have obtained without following the recommended practices?

9(a) What is the total area you have planted in this crop? _____

(b) How much of that is under newly recommended practices? _____

10. How would you rate the usefulness of the agriculture extension program?

- 1 Extremely useful
- 2 Quite useful
- 3 Useful
- 4 Not of any real use
- 5 The recommended practices are wrong

Crop Cutting Section

11. Code number for identifying the plot and the dried weight of produce from plot: Code _____ Weight _____

Figure 2.4 (Contd.)

12. Green weight of produce from plot _____

13. Was crop sown pure mixture

Field Investigator

Date

Central Group would also oversee the organized transfer of research experience to the units and synthesize the units' findings for nationwide policymaking.

Much can be said in favor of the RORSU design for rural development M & E systems, if one accepts the assumption that data collection and analysis should be carried out as concisely as possible for management purposes. There are weaknesses in RORSU's framework, however, that might inhibit the process of locating trouble spots in project implementation. First, RORSU ties M & E activities largely to the regular project reporting system, without designing a reporting system tailored to the evaluation and tracking efforts themselves, or to the additional data generation mechanisms they may require. While RORSU might argue that such arrangements must be project-specific anyway, it is hard to comprehend why nearly five years of formal review operations in the sector have not yielded even a tentative set of indicators pertaining to those parts of the management information systems that have regularly been bottlenecked. Nor has RORSU highlighted the kinds of information that have been reporting smoothly, obviating the need for a separate monitoring system.

Secondly, RORSU's approach is inconsistent on the matter of relations between the monitoring and evaluation functions themselves. Monitoring is said to concentrate on "what" is or is not functioning as planned in a project, whereas evaluation is charged with determining "why" particular obstacles or successes occur. Yet both monitoring and evaluation are supposed to yield recommendations on how to identify and

alleviate in-course problems. How can monitors offer perceptive advice on shifting project courses by focusing on only "what" is going wrong without considering "why" problems have occurred? If monitoring is charged with comparing actual to planned inputs and outputs, and evaluation is charged with comparing actual to planned effects and impacts, of what does the supposed focus of the two on "different" activities consist, other than their timing in the project course? The argument implicit in this book is that a basic evaluation system, planned along with the project reporting systems, and clarifying the information bases of each, would provide a more streamlined and effective means of project control than institutionally (or nominally) separating functions called "monitoring" and "evaluation."

There is a danger too in the sharp divide between "minimum information" M & E surveys and longer studies proposed by RORSU, to the extent that crucial information may slip between the two. For example, Cernea seems so concerned with expeditious evaluation that his draft survey for the exercise omits the specific question of why contact farmers have cropped the way they have. The questionnaire might well be strengthened by a direct interrogation on the farmers' rationales for their actual cropping patterns. And though Cernea proposes longer evaluation studies on farm practices and village social networks, his Working Paper is vague on the linking of survey questions to subsequent studies.

Again with regard to the questionnaires, it is unclear why the monitoring survey, but not the evaluation survey, asks the farmer for

information regarding the spread of practices from himself to other villagers, or why the evaluation survey has no open-ended question on the utility of the extension service. Both surveys ask for an overall judgment of the extension program without specifically asking the farmers for their opinions on its components; for example: how well they understood the recommendations, the amount of direct help they received from the extension worker, the value of group meetings, etc. The indicators chosen are not so much problematic as insufficiently specified; some of them require sub-indicators for interpretation. For example, should "quality of visits" be determined strictly by the number of desired practices adopted, or correlated with the availability of fertilizers, credit, pesticides, and high-yield crop varieties? Should the village extension workers be evaluated on how well they test after training, on their monitored performance with farmers, or on a combination of these factors? And while one can certainly appreciate the difficulty of establishing formal control groups in the sector's evaluation designs, it would seem possible to identify some useful comparison group from other projects in states not concerned with using extension services as a focus for development. Although cross-project comparisons raise their own peculiar problems, some sense of the extension mechanism as a variable in the overall area development could be gained from such an exercise.

For reasons apparently having as much to do with project implementation as with the M & E system, the Cernea proposal for aiding assessment of agricultural extension projects has not been adopted.

Similarly, Guido Deboeck's "Systems for Monitoring and Evaluating Nutritional Interventions," although well received within the Bank, and in the target country, Brazil, has not been activated due to project delays. RORSU has meanwhile concentrated on serving as a convener of forums for evaluation officers and project managers in different regions, and on acting as coordinator of their discussions.

The Case of Northern Nigeria: Where Evaluation Proved Its Value in Rural Development

The most notable relative success of an M & E component in the rural development sector has been that of the Agricultural Projects Monitoring, Evaluation and Planning Unit (APMEPU) in Northern Nigeria. Set up as a department of the Nigerian Ministry of Agriculture and Rural Development in 1975, the Unit was originally charged with creating an evaluation system for the initial three (which have grown to seven) area development projects focused on improving rainfed farming. There is little doubt that the promising performance of APMEPU has been due partly to its substantial funding. Combined funding for the three original M & E units within the state-based projects totaled a very considerable \$2.7 million, with an additional \$7.2 million allotted for APMEPU itself under one of the projects.

APMEPU comprises two sections. One reviews and oversees financial records. The other has developed the evaluation system, providing Federal and State government and other institutions with relevant economic and social data on project performance. In addition, the latter section provides technical direction to the project-level

evaluation units, to ensure adequate data collection for project-specific monitoring, as well as over-arching analysis. It has also served as trouble-shooter on project procurement, agricultural extension, and training.

The Evaluation Section of APMEPU primarily concentrated upon initial village listing and both baseline and punchline surveys in all component projects. From a baseline survey (in the first three), a 6% village sample was drawn, in turn determining a household sample for the punchline surveys. Populations were stratified by village type and size, and produced good geographical coverage on stock variables (household size and composition, occupation, landholding, cropping patterns, market channels, credit position, asset ownership, etc.). Mainline surveys of flow variables, later simplified to punchline, combined income, expenditure, and farm management data. Data gathering was undertaken twice weekly on a sample 623 households, and alternate methods were used to test manageability and acceptability of each. Although major difficulties were encountered in Nigeria both with keypunching and computer facilities for programming the data analysis, these were eventually resolved, although masses of data had to be brought to Washington for final processing and analysis. In addition, some 20 discretionary studies were also carried out by the Evaluation Section, based on its data findings. These studies covered such topics as fertilizer requirements for annual use in the three original projects, effectiveness of alternative extension methods, effects of tractor-hire practices in the project area, and price support policies

in the region. Whilst project management, often for political reasons, paid little attention to these succinct, pointed studies, they constitute a replicable model of problem-orientation, investigation, and solution for both rural and urban contexts.

What makes the APMEPU case so noteworthy is its commitment not only to careful selection and verification of indicators of rainfed farm development, but also to understanding the tactics of farmers in their planting, marketing, and labor utilization. Data thus collected for annual tabulation were used both for policymaking issues, such as the relative importance of fertilizer deliveries to different areas, and for documenting the rationally economic bases for farmers' activities. The data could be used to answer the questions why planters preferred to intercrop on single stands with resulting efficiency; why early planting was shunned due to lower prices early in the season; and why inputs such as fertilizers and spraying were used discriminatingly to support stands yielding a high and stable income. APMEPU envisioned its role as one of testing the assumptions of project appraisers regarding both the existing systems of farm management and the appropriateness of recommended new technologies for the area. Nearly five years of on-site investigation and Washington-based analysis have established that the appraisal reports misunderstood to some extent both the actual farm practices of Northern Nigeria and the technical possibilities of increasing productivity and marketability of commercial crops. Although the cost of such information has certainly been high, Nigeria (and the Bank) now have possession of data on patterns of crop growth under a

variety of technical, ecological, and economic conditions that can be used to improve project planning considerably and thus to ensure better use of resources in the future. Already several additional interventions are being programmed for the next five years. These are based on the new information, coupled with findings from a Bank research project which further assesses constraints on the adoption of new technology at the farm level.

Clearly, not many projects can expect to benefit from the funds, staffing, and analytical capability specific to the Nigeria case. The success of APMEPU shows, however, that where evaluation efforts are properly supported by the borrower and the Bank, the results not only justify the expenditure of resources, but promise to enrich the benefits derived from subsequent projects.

APMEPU went beyond Michael Cernea's recommendations for evaluating agricultural extension services by conducting a survey in 1977 which collated interviews of agents on their recommendations to farmers with farmers' reactions to the recommendations and their attitudes toward farm technology in general. The results showed that agents retained only 15% of their project-controlled training; at best visited farmers only at random; and gave out little information that non-project farmers did not have. Although farmers claimed to prefer face-to-face contact with agents, they apparently paid more attention to project radio broadcasts than to agents as a source of farming information. Despite frequent coordination difficulties with project management, APMEPU and its satellite units have continued to collect and

process timely information on the outcomes of a variety of project components.

OED's View of Rural Sector Evaluation Efforts

Finally, mention must be made of the Operations Evaluation Department's first review of built-in monitoring and evaluation in the agricultural and rural development sector, dated October 4, 1977. ^{1/} That review examined six projects, all in Sub-Saharan Africa (in Nigeria, Lesotho, Senegal, Upper Volta, Mali, and Malawi). It uncovered the same complaints registered in nearly all other interdepartmental Bank reviews of M & E: data collection exceeds processing and evaluation capacity; high cost of evaluation activities; evaluation units' failure to answer the real needs of project management for decisive information affecting policy; high staff turnover, etc. Nevertheless, OED concluded that:

Results obtained so far through M & E seem to have largely justified the costs involved. All units but one were felt to have contributed to improving ongoing projects, provoked beneficial adjustments of second phase projects, or provided valuable information about the farmers' constraints to development.

Yet OED itself failed to specify its own criteria for deciding where and when evaluation costs were justified by results, let alone deciding what the principal substantive problems of M & E units were. Although the report cites extremely significant shortcomings on the Bank's and the project's parts in handling M & E components, it does not

^{1/} OED, "Built-In Project Monitoring and Evaluation: First Review," Report No.1758, October 14, 1977.

grade these liabilities according to a demonstrated scale of cost-benefit accounting or cost effectiveness. OED seems to have evaluated the M & E units on an overall "worth it/not worth it" basis, without providing its Bank audience an account of the interim steps taken to reach its conclusions.

OED did, however, go to the heart of the evaluation issue in pointing out how rarely a direct relationship can be found between M & E findings and project management decisions. The report concluded that the evaluation units are liable to generate useful but unused information, repeat errors, or carry out unnecessary exercises until the Bank deepens its own capacity to absorb and act upon evaluation findings. Subsequent chapters detail the process by which the Bank has learned the lessons of evaluation and gradually begun applying them in the even more complex environment of the urban development sector.

PART II

Chapter 3:

THE DESIGN OF URBAN PROJECTS: PROGRESSIVE DEVELOPMENT AND AFFORDABILITY

The Progressive Development Model

When the World Bank began to commit vast resources to the development of urban housing in Third World countries, the design of its projects adhered with increasing explicitness to what is called the progressive development model. Progressive development differs from traditional development schemes in several important respects. World Bank projects have been of two basic types: upgrading, and sites and services schemes. The progressive development approach assumes that it is better to take advantage of existing housing stock within a city, whatever its condition, than to demolish and replace it. Upgrading schemes provide essential infrastructure and secure tenure arrangement; and then extend credit, materials and technical assistance to dwellers in existing urban slums for the improvement of their own dwellings. Sites and services schemes provide contractor-built utility infrastructure and core houses, leaving project participants in a position to complete their own new dwellings, again with provisions for materials loans and technical assistance.

Both types of progressive development strategy attempt to incorporate as much flexibility as possible into specific project designs, giving participant families a maximum of options as to period and method of construction, materials, and sources of labor. A key

design feature of both upgrading of sites and sites and services projects has been provision for participants to use self-help and mutual-help labor. The flexibility of such programs was intended to enable families to adjust the pace of house construction to their fluctuating incomes, and to use their own labor (or exchange labor with others) to reduce "out-of-pocket" expenditures. All these features of progressive development practice were intended to enhance the accessibility of improved housing options to low-income segments of the urban population, and even to affect the quality of social existence within poor areas. It was expected that the encouragement of self-help and mutual-help house construction would aid in promoting a sense of civic responsibility and community solidarity previously lacking in impoverished neighborhoods. Though there are still those who remain to be convinced, and the inevitable problems of project implementation and the occasional "problem projects" fuel these doubts, the results of these efforts to improve the quality of urban life on a mass scale have so far been encouraging. The successes have been numerous enough and dramatic enough to cause project administrators and designers to feel that the progressive development model has proved its validity as a strategy for coping with ever-burgeoning urban populations in the Third World. The main persisting theoretical objections to the progressive development approach, and technical indicators of its viability are summarized in Appendix I.

This chapter details the results of more than five years' experience in the effort to make improved urban housing options

affordable to the poorest segments of urban populations. For an understanding of the evaluation programs developed to derive the lessons of urban housing project experience, it is necessary to become familiar with the problems posed by the design of the projects themselves. For this reason, the current three chapters treat in detail the structure of the World Bank's major urban housing projects and evaluation methods and design decisions they imply. Part III of this book describes the actual evolution of the World Bank's urban sector evaluation program that has yielded so many of the findings presented here.

Designing Affordable Projects

The design principles followed in the World Bank's urban projects center on the practice of delivering "packages" consisting of tenure, services, and assistance in house construction, either in new subdivisions (sites and services) or in existing squatter areas (upgrading). The design of these projects seeks to ensure that beneficiaries are neither constrained against spending what they want and are able to spend, nor induced to spend beyond their capacity and willingness. These considerations define the concern for affordability. Closely related are the issues of accessibility and replicability. Accessibility is achieved when families within the prescribed range of incomes are those that actually gain entrance to the project. Replicability refers to the goal of having project costs paid for by the beneficiaries, with little or no subsidy involved.

There has been a tendency to treat these three terms -- affordability, accessibility and replicability -- as synonymous, but

clearly they are not. The aim of improving project design is that these three objectives will converge, but there is no assurance that they will. If a project is not affordable, it can either be replicable (by the entry of high-income families) or accessible (through subsidy of lower-income targeted families), but not both. On the other hand, an affordable project may not be replicable (if the cost recovery mechanism fails) or accessible (if, for example, selection procedures favor higher-income families over members of the target population. This chapter focuses on the concept and determination of affordability as a design factor. Replicability and accessibility are considered in detail in the following chapter.

The Theory Behind Affordability

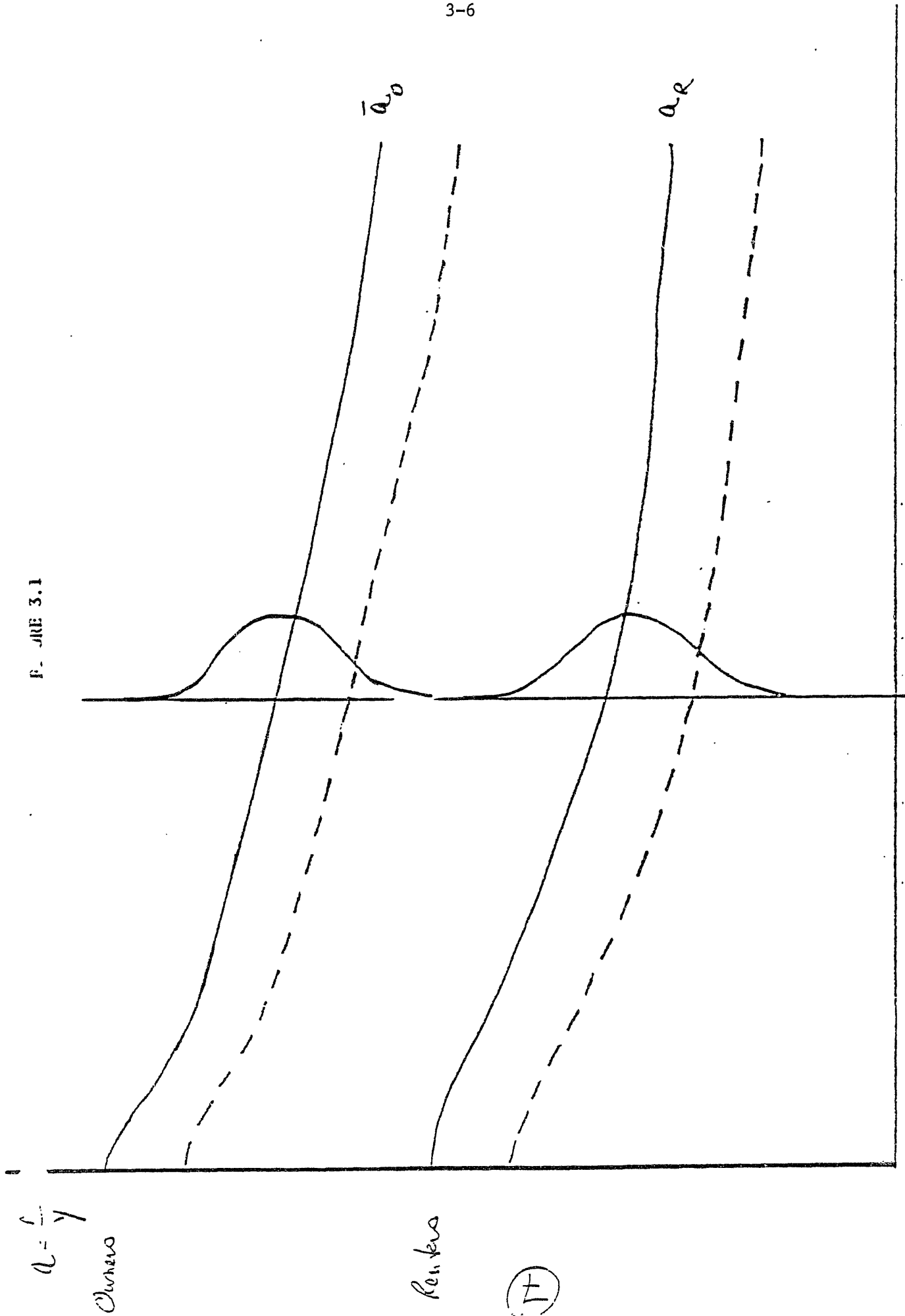
When we think through the relationships among variables, and between project design and evaluation, demand theory leads us to posit a general relationship between housing expenditure and income. Whether we have gathered information or merely made assumptions about the income (Y) of any population, it will be a distribution, or vector, as are the other variables to be considered. It is necessary to reflect on the formula, $a = r/y$, in which "y" = monthly income; "r" = monthly rental payments or occupancy costs; and "a" = the average propensity to consume housing services, which is a function of "y". It is important to realize that "a" is different for renters than for owners; and that, as a function of "y", "a", at every value of y_i , has a mean a_i , and a standard deviation O_i . This is because "a" is in addition a function of other variables, such as relative prices, (and therefore level of

development national income level), city size, family size, sex of household head, family assets, and certain characteristics of the extended family. The essence of the situation is represented in Figure 1, where the solid lines refer to the loci of means, and the broken ones correspond to a general notion of the proportion of the population that housing policies should try to reach.

The tasks of project designing begin with the determination of income distribution by measurement, estimation, or inference from comparable populations.

In essence, we want housing programs and projects that "reach" as far down as possible into the income distribution. We also want replicability, and thus good cost recovery performance. For this reason we have admission criteria: we want to admit "successes," and exclude "failures". However, this is not a simple task. We may err in several ways: By pushing the lower income cut-off too low, we may admit households that will fail -- at high costs to individual projects and the perceived credibility of the approach. If we keep the lower income cut-off too high we may unwittingly reject poor households that would have succeeded, at a high direct cost to distributional objectives. Towards the upper end of the target distribution we may want to encourage some participation at higher prices to produce desirable cross-subsidy and market interactions. However, we so far lack information on the costs and benefits of this approach, and so have no inkling of optimal or satisfactory proportions.

F. JRE 3.1



$a = \frac{l}{y}$
Owners

Renters
 $(+)$

y

At both ends of the spectrum, but particularly at the lower, it is necessary constantly to apply improving affordability criteria, and judgment. The need for sound judgment is acute, because we are dealing with a statistical relation--with considerable scatter around the curves--and not a natural law. Affordability criteria should thus be regarded as a decision aid, which must eventually be made relatively inexpensive to apply, not as a source of decision-making rules.

Findings to date indicate that if the lowest income groups are to be reached, total income must be carefully estimated and public costs of a project kept to a minimum. These measures imply the relatively low standards of minimal core house construction so far typical of progressive development practice. These steps insure against screening out large numbers of families which, though poor, can afford housing options. At the same time, project designers must be careful not to constrain unduly the behavior of households that have both the desire and capacity to do more than the minimum. Scatter diagrams (Figures 3.2 - 3.5 show that participants' behavior varies greatly. What they record is the interaction of a number of forces that sometimes conflict and sometimes reinforce each other. These include variable incomes, variable investment plans and variable timing of house construction and variable investment plans. Inconsistencies are probably accentuated in formative areas, where one can see shacks juxtaposed with completed family dwellings, and large houses containing rental accommodations for several.

In the long run, it will probably prove undesirable to encourage a wide diversity of participants in projects. Yet a tolerance of such diversity can be a means of moving projects toward the fulfillment of their major objectives. This approach calls for considerable flexibility in credit programs and other project components so that they can address a wide range of the population's needs and capacities.

Given that "a" varies with so many factors, and broadly with some, an average value of "a" is at best a very rough guide to the other factors in project design. An average figure will pose problems both for households with a low "a" whose income is adequate, and for those whose high "a" would offset a low income, in an unrestricted situation.

The best possibility, it if can be done, is to generate an expected "rent" profile for a given population's income distribution. Projects should then be designed to eliminate only the bottom (say, 5 to 20 percent) of this profile, where all factors are taken into consideration. What the data show, particularly at low levels of income, tends to be obscured by the use of averages: some families spend next to nothing on housing, others spend a great deal. As long as incomes make possible the amortization of costs, the full range should be accommodated. This premise argues in favor of projects with very low entry costs, but the opportunity to add substantial amounts of investment. This is the opposite of what most designs for low income housing offer, but it is precisely what the private markets provide, albeit usually in an illegal or quasi-legal fashion.

Figure 3.2

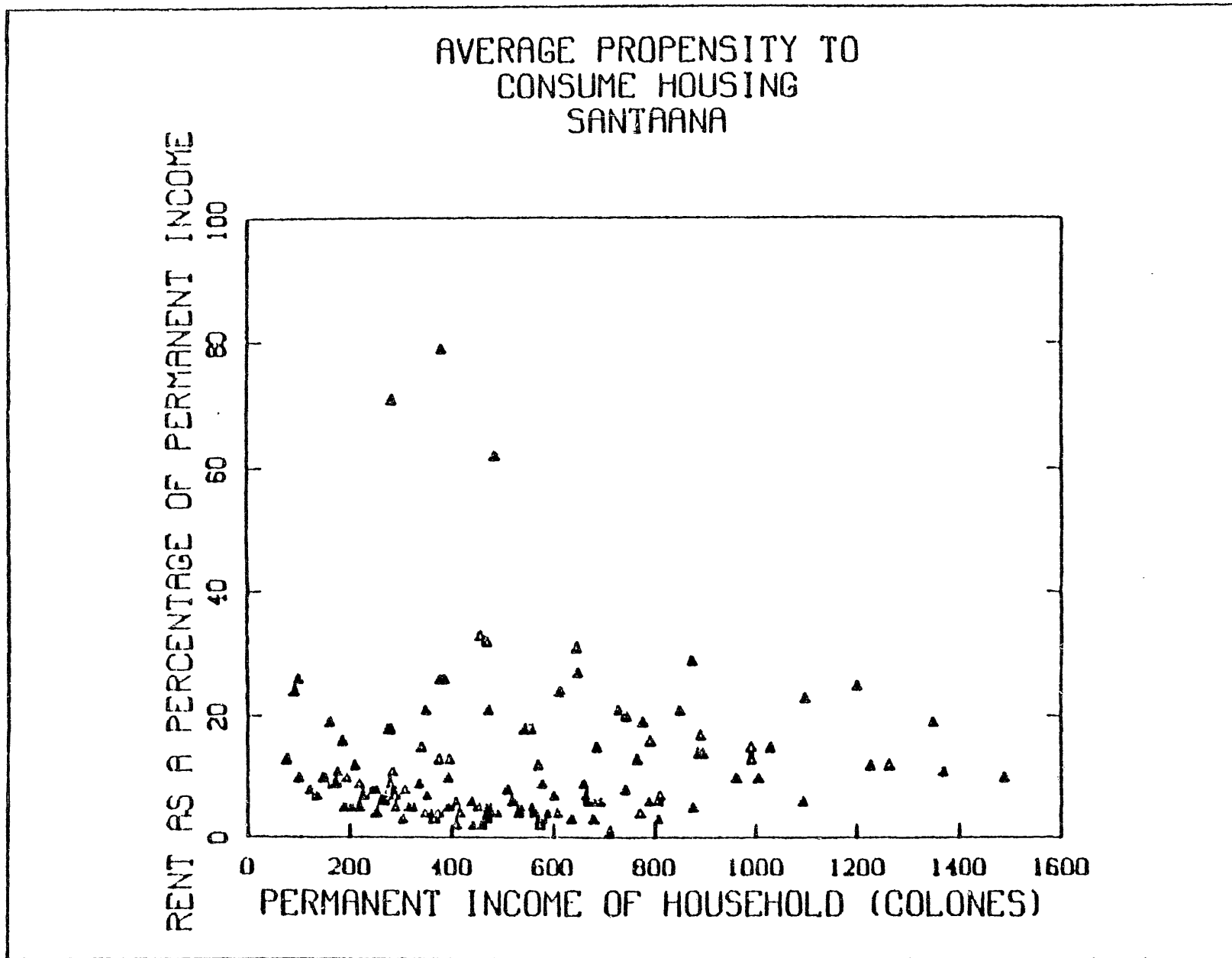


Figure 3.3

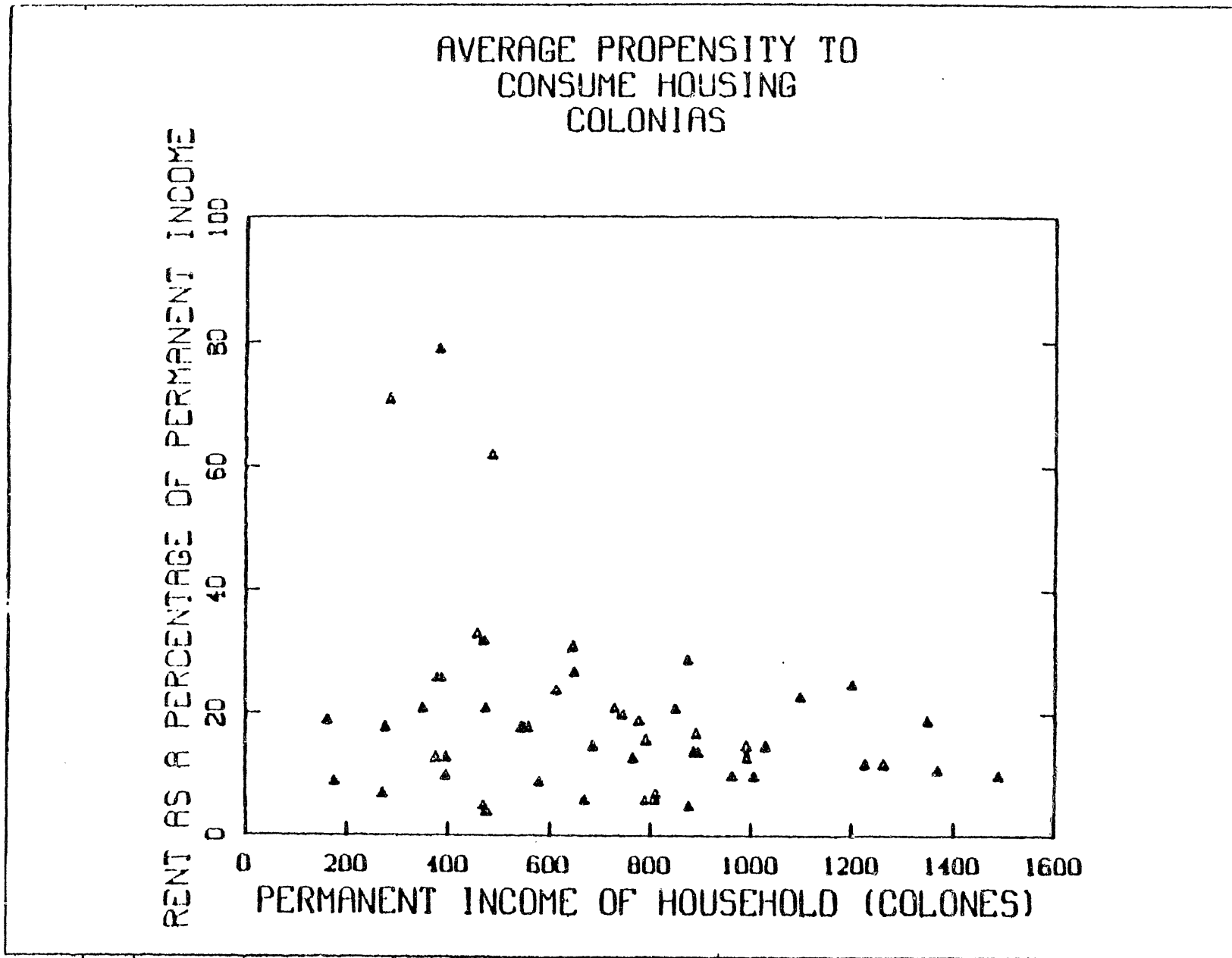


Figure 3.4

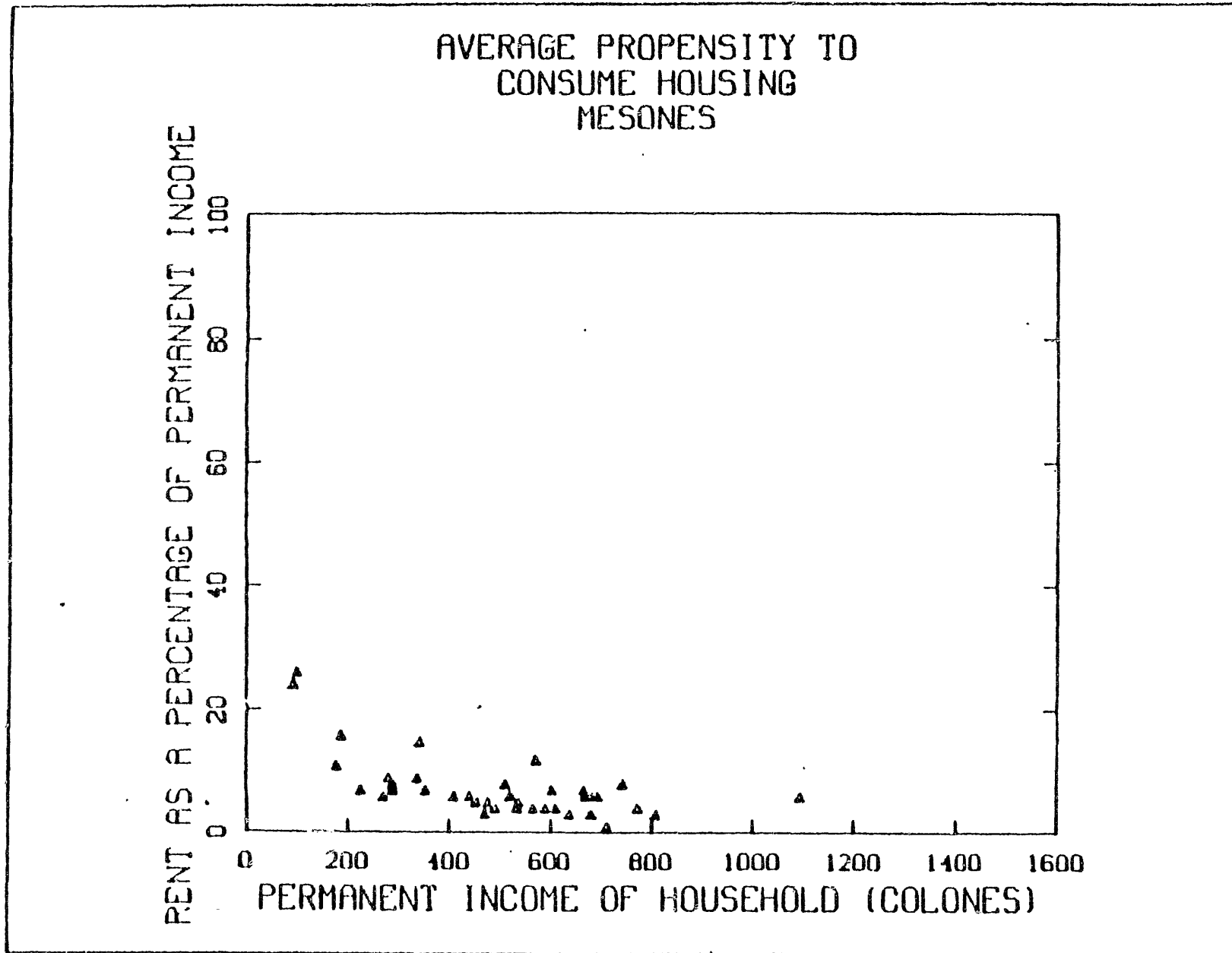
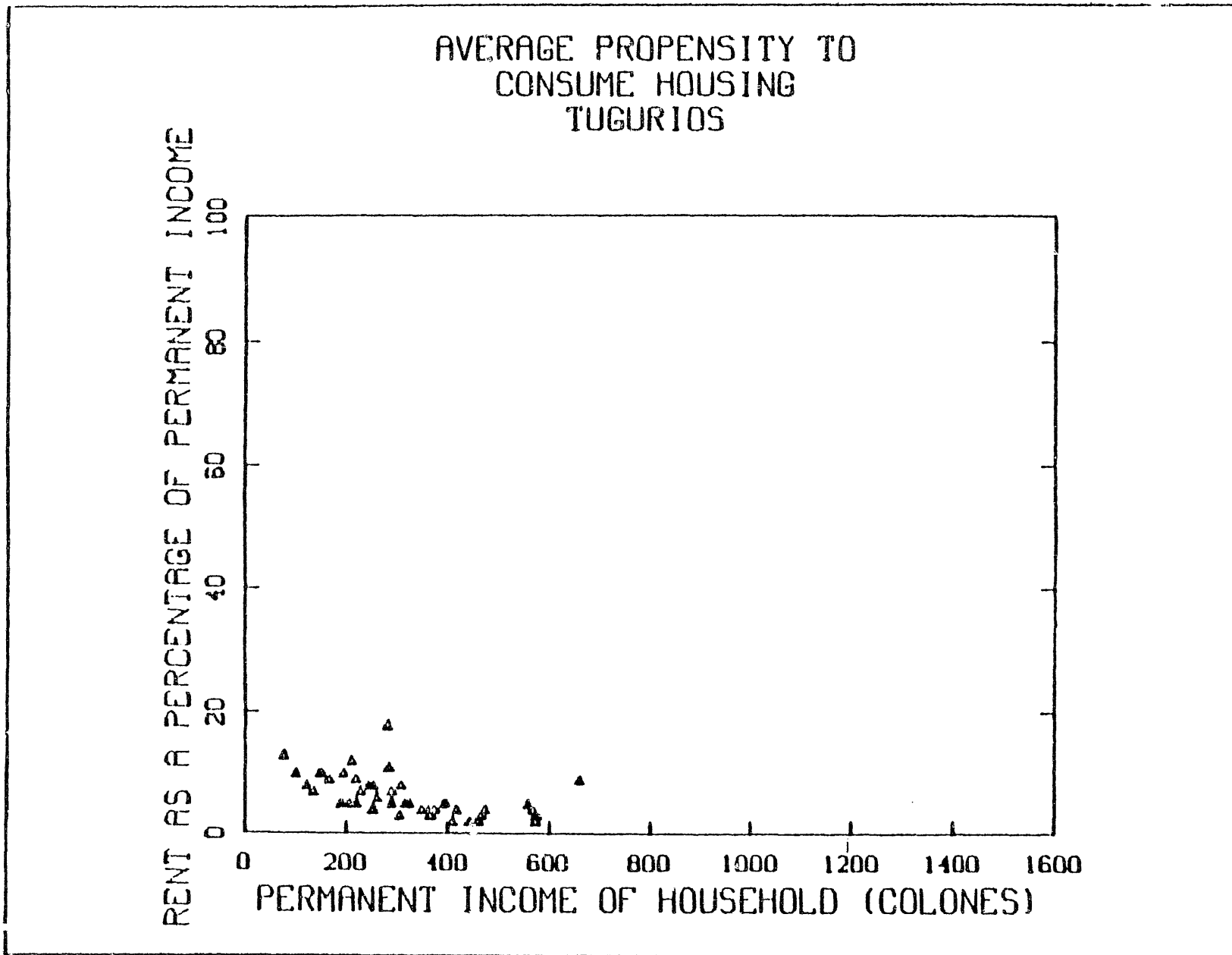


Figure 3.5



It is important to note the differences between renters and owners, among households. Families should never be consigned to live permanently with their relative priorities of the moment, nor forced to adopt new priorities (different "a"s) precipitously. For example, in a project designed for a target population consisting of large number of renting families with representative incomes, but low (ex-ante) "a"s, the use of standards based on owners' behavior (i.e., higher implied "a"s) will result in a large percentage of these families being designed out of the project.

There are basically two ways of avoiding this outcome. One is to design potentially excluded families in by setting base standards very low. The other is to design them in, as continuing renters, by setting higher standards but providing specific rental components or incentives.

The Evaluation of Affordability

For more than a decade, the World Bank has been engaged in efforts to design and implement workable solutions to the shelter problems of the low income families that have been swelling the populations of Third World cities. Again, the working definition of affordability used in appraisal reports has been as follows: a certain level of urban services is affordable to a low-income beneficiary household if the amount the household is willing and able to pay for shelter-related expenses is sufficient to cover the monthly cost of services. The definition may be written symbolically. If "a" represents the average propensity to consume housing, the proportion of

monthly income a household is willing to spend on it, then a project housing unit with service level "j" and monthly costs " C_j " is estimated to be affordable down the "i"th percentile for the income distribution with monthly income " Y_i ", if the following is true: $aY_i \geq C_j$. ^{1/}

As an example, consider a project whose costs are such that, if it is to pay for itself, beneficiaries must be charged \$100 per month. If all households are assumed to be willing and able to set aside 20 percent of their income on housing ($a = 20$), then a household must earn at least \$500 per month to afford project housing. If the income distribution is as follows:

<u>Percentile of Population</u>	<u>Y = Monthly Income</u>	
14	Y_{14}	\$490
15	Y_{15}	\$500
16	Y_{16}	\$508

then the project is affordable down to the 15th percentile of the population.

The Bank's interest in examining affordability has hinged upon avoiding projects that are too expensive for target groups, thus

^{1/} This will be recognized as the same relationship as that introduced above. However, C (cost) has been substituted for r (rent). In the estimation of "a" we must typically rely on measures of monthly rent, whether actual or implicit. When using the resulting estimates of "a" to assess affordability, we are concerned with design costs. To fill the bill, the cost concept used must be (fully amortized) occupancy costs.

emphasizing keeping standards and costs down. This explains why the affordability criterion has been written: cost must be equal to or less than the affordable proportion of income, and why both operational and evaluation programs of the Bank have stressed the harmful effects of designing projects that target groups cannot afford.

The estimated value for "a" is a crucial element in calculating affordability. Until very recently rigorous estimates for developing countries did not exist because the evaluation and research programs designed to inform the Bank's policy decisions had not been in operation long enough to generate appropriate data bases. Policy formulation and project design relied upon judgments concerning the stability of relevant parameters across geographical boundaries and income levels.

Grimes was one of the first to collect information for developing countries.^{1/} Using aggregate data, he concluded that:

the average percentage of household expenditure devoted to housing ... falls into a fairly narrow range (across cities), from 11.7 percent in Kingston to about 20 percent in Mexico City and Seoul.

However, in looking at some rough and strictly non-comparable government micro-level data for five countries, Grimes also found that:

The fragmentary data available on expenditure by various income groups ... indicate a general tendency for

^{1/} Grimes (Ref.)

housing expenditure by low-income households to claim a higher than average share ... As income rises and other demands are met, the share of income devoted to housing may remain constant or fall. 1/

By the second year of the evaluation program, it had been determined that, not only was the progressive development model generally viable (in the sense that housing was being completed and occupied, or up-graded, according to the case), but the first generation projects being supported by the Bank were also affordable in terms consistent with Grimes's findings. Based on initial, detailed evaluation of one sub-project, the El Salvador project appeared to be affordable down to about the 20th percentile of the income distribution. 2/ Despite measurement difficulties, this result seemed to be borne out as well by the Zambian and Senegal evaluations, and to be broadly confirmed by other project experiences up to that time.

These generally encouraging results were subject to more than one interpretation among the project managers, researchers, and institutions involved. For example, the precise results in the case of El Salvador showed that, assuming an average propensity to consume of 20 percent, projects costs were affordable to those with incomes down to the 17th percentile. However, if a crude estimate of additional private costs was included, it appeared that total housing-related costs might be affordable only down to the 25th percentile of income distribution. The questions raised by such findings pointed to the conclusion that affordability is a matter of greater complexity than had been assumed.

1/ Ibid., pp. 65-67.
2/ (Ref.)

By the time an adequate data base had been accumulated, it had been found that the values of "a" assumed in the first generation of Bank-supported urban projects had been distributed about the upper end of the narrow range described by Grimes. An internal analysis of 36 projects approved from 1972 through 1979 recorded extreme values for "a" of 8 and 50 percent. However, the bulk of the observations were approximately normally distributed about a mean of 20 percent, within a range of 12 to 33 percent.

What is striking about these figures is the large variation in a parameter previously considered by most to be quite stable. Several factors apparently contributed to this outcome. However, it is essentially a matter of definitional and quantification problems involved in measuring both income and housing expenditures. The evaluation program adopted several complementary goals in its analysis of affordability. The primary goal remained that of obtaining more accurate estimates of "a". Several questions regarding this parameter must be answered in order to gauge its value as an operational tool. Is the value of "a" constant (or close to it) across various income groups of the target population? In most projects, Bank staff have assumed that "a" is fixed. Some projects (Ivory Coast, El Salvador II, Thailand, Botswana II, and Brazil) tacitly assumed that the proportion rises with income. The Bolivia project apparently assumed the opposite tendency. The divergent assumptions naturally lead to the next question: If "a" is not constant, how does it vary with income? How sensitive is "a" to the definitions of housing expenditures and income?

The next section of this paper attempts to deal with these questions by recounting the investigation of determinants of housing consumption in El Salvador.

Determinants of Housing Consumption: General Principles
and the Case of El Salvador

The problems in estimating average propensity to consumer housing are complicated by the non-homogenous nature of housing, and by the uncertainty and variability of personal income in poor countries. The problems dealt with here concern definition and measurement of the income (Y) and housing expenditures (C) of a prescribed segment of a market or population. Since these figures are used to estimate $a = C/Y$, errors affect the estimate. (Other kinds of measurement problems, treated later, arise with attempts to use an estimated "a" and assumed or measured Y for a target group, to establish design criteria for project cost.)

The data for this analysis have been collected in El Salvador as part of the evaluation program. They consist of responses to extensive socio-economic questionnaires applied in 1977, 1979, and 1980 to samples of approximately 260 households in each of two cities, Santa Ana and Sonsonate (the number of families varied with attrition over time). These comprised the control groups for the evaluation, that is, families continuing to live in one of three types of housing: rental rooms in mesones, quasi-legal dwellings in colonias ilegales, and illegal dwellings in tugurios. In colonias ilegales, legal owners sell plots to legal purchasers who then build houses on them. These subdivisions, however, do not meet subdivision codes, and the homes

built there may not meet building codes. Tugurios consist of groups of shacks and hovels thrown together in rights of way, arroyos, and other undesirable locations.

It appears that estimates derived from data on these families will prove superior to those previously available to guide project design. Measurement of income and housing expenditures has been more complete and accurate in this case than in most other comparable surveys. The richness of the survey enables researchers to control for several socio-economic characteristics in the analysis. In particular, it will be possible to estimate the differences in the effects of various components of total income, which may be as important as the wage income of the household head in determining the purchasing power of the poor. Further, the nature of the panel data permits estimates to be made using an approximation of permanent income not previously applied.

To minimize the risk of measurement errors, definitional ambiguities must be removed. In addition to the decision whether to attempt to develop permanent income measures, questions inevitably arise as to whose incomes and what sources to include. In early research efforts, a developed-country bias crept in, resulting in a tendency to think only, or primarily, of the household head's wage income, since that was judged to be the most stable component of total household income. A number of studies have since lead unambiguously to the conclusion that this assumption brings about serious underestimations.

It is now quite clear that total household income is the figure to be used. 1/

Evaluation program research further suggests that all income sources--earned income as well as unearned (including transfers)--should be taken into account. There is still considerable resistance to this idea, however, and most estimates to date have been made on the basis of earned income sources only. The disputed practice is partly a matter of definition, and partly of the added difficulty of getting reliable information on gifts and transfers. One argument has been that unearned income should not be considered when estimating affordability, because such income is subject to the whims and tribulations of others, and is therefore unreliable. Recent studies have indicated, however, that transfers, particularly within the extended family, are an important and stable component of household income among low-income urban families in developing countries. 2/ Moreover, it appears that transfers may be allocated in a compensatory manner, such that the most impoverished families have their basic needs met, as demonstrated by the research of Kaufmann and Lindauer. 3/ So, in fact, total income may be less variable than earned income, not more.

1/ Dani Kaufmann, "Household Income Formation and Expenditures Behavior: A Summary of Issues, Findings and Research Prospects," Urban and Regional Economics Division, Development Economics Department, World Bank, 1981.

2/ Ibid.

3/ Dani Kaufmann and David Lindauer, "Basic Needs, Inter-Household Transfers and Extended Family," Urban and Regional Report No. 80-15, World Bank, 1980.

At this stage it is clear that the most appropriate gauge of income is total household income --including the incomes of all family members from all sources -- adjusted or averaged in so far as possible to approach a permanent income concept. Even if this definition can be agreed upon, however, formidable measurement problems persist. The first of these is that the available data sets are typically one-time household surveys which provide bases neither for averaging nor for adjustment, let alone for purposes of estimating permanent income. The World Bank Evaluation Program enjoys two tremendous planned advantages in this respect. The first is that surveys are repeated to give at least three readings at one to two year intervals. The second is that an unprecedented survey in the Philippines is following incomes and expenditures of a small sample of households over a three year period. This work is generating information and insights that will eventually make it easier to cope with a broad range of measurement and estimation problems.

The remaining measurement problems relate to capturing accurately all the income sources for all persons within a household. These problems include failure or refusal (by the "counters") to record all the sources; difficulties attendant on collecting unearned income information, such as lack of recall and concealment; and complications caused by such social arrangements as polygamy. It is possible to list several likely causes of data unreliability. One is the impossibility of accounting for fluctuations in incomes in one-time surveys. Another is the informants' possible ignorance concerning total income sources,

whether due to poor record-keeping, secrecy of other family members, or inability to reckon the effects of such activities as self-help construction or the families' raising of its own food. Then there is the possibility of outright fabrication of figures, perhaps to meet project criteria, or to avoid purveying information that might reach the tax collectors, or simply to get rid of the interviewers. Such problems have consistently plagued analysts attempting to obtain accurate estimates of income. Though the difficulties are real, they are often exaggerated by researchers uncomfortable with quantification. Much is now understood about all aspects of sampling, surveying and cross-checking; and experience has shown that acceptable results can be obtained with careful, cost-effective application of this knowledge.

Measuring Housing Expenditures

In gauging or estimating housing expenditures, or the value of housing services consumed during a year, the first problem is to be clear about what is being measured. This is a relatively straightforward matter where renters are concerned. The rental charge is usually unambiguous and consistent, although there are practices (particularly where rent controls are in vogue) such as "key money", surreptitious payments and excess charges for certain services, that make the calculation more elaborate. The task is somewhat more difficult with respect to owners, since there is seldom any summary measure of housing expenditures over a given period, such as a month. For low-income households, self-help in house construction complicates the issues further, as one must also place a value on non-purchased inputs,

such as family labor and, in the case of developing countries, scavenged materials. Simply truncating samples so as to exclude owners would seriously distort estimates of "a", since owners are deemed willing and able to spend considerably more on housing than renters.

There are two ways of approaching this problem. One is to bring all costs down to an estimate of monthly "occupancy" cost. For a host of reasons, this figure is difficult to estimate accurately. Thus a second, easier method is more often relied upon. This approach entails asking the owner two questions: What would you sell your house for? and What would you rent it for? (It is assumed that, if inflation exists, the respondent is equally aware of it in answering both questions, and that money illusion therefore does not affect the results.) A ratio can then be made using these two values:

$$R/V \text{ Ratio} = \frac{\text{Monthly Rental Value}}{\text{Sales Value}} \times 1000$$

Typical values for this ratio, as measured in markets where like properties are both sold and rented, are on the order of 5 to 10 -- implying real rates of return on the order of 6 to 12 percent per annum---although substantially lower values are recorded in upper-income residential areas where rentals are relatively infrequent and the expectation of eventual capital gain on resale is a dominant market feature. El Salvador figures (Table 1) are generally consistent with

Table 3.1
 RENT-VALUE RATIOS BY INCOME DECILE
 SANTA ANA AND SONSONATE OWNERS, 1980

Decile	Santa Ana		Sonsonate	
	Mean $\frac{1}{\text{Income}}$	R/V Ratio	Mean $\frac{1}{\text{Income}}$	R/V Ratio
0-10	151	13.5	-	-
10-20	273	25.3	-	-
20-30	355	11.3	219	4.1
40-50	527	8.5	325	8.5
50-60	696	9.2	390	6.9
60-70	820	6.7	471	13.7
70-80	917	7.1	546	6.9
80-90	1094	9.0	772	4.8
90-100	1582	6.0	1025	4.8

$\frac{1}{\text{Income}}$ Colones per month.

these norms, as well as with the results of similar recent studies in other developing countries.

The larger than expected K/V ratios for the bottom of the income scale in Santa Ana appear to be attributable entirely to tugurio dwellers since when they are removed, the values for colonia ilegal dwellers range only from 4.1 to 8.2, whereas those for tugurio dwellers range from 14 to 30 percent in the bottom four deciles. One may hypothesize, then, that tugurio dwellers tend systematically to underestimate and colonia ilegal dwellers perhaps to overestimate the resale value of their dwellings relative to their rental values. This is not hard to imagine. Few tugurio dwellers are likely to think in terms of resale of their precariously situated huts of cardboard and wattle. They will first be evicted or their homes washed away in the next heavy rains; all bits of value will be carried away to the next location. By contrast, lots in colonias ilegales are usually well short of their full development. Lots are large. Progressive development is typically applied. The period of development is likely to be extended compared to that in sites and services areas. It would not be not be surprising, therefore, if colonia ilegal dwellers tended to think in terms of a higher future value when pondering resale.

Housing Consumption in El Salvador ^{1/}

Calculations of "a" for renters in mesones in each of the three survey years for Santa Ana and Sonsonate indicate that even the

^{1/} The material presented here is a selection of findings from Emmanuel Jimenez and Douglas Keare [Ref.]

simplest reckonings of the average propensity to consume housing, made with carefully generated data, are far from straightforward. Because wage income of the household head accounts on average for only about 63 percent of total household income, the denominator of "a" will be seriously affected by the definition of income chosen. A crude average over three years gives the following results:

Average Propensity to Consume Housing

<u>Out of</u>	<u>Santa Ana</u>	<u>Sonsonate</u>
Wage income of household head	.20	.15
Total household income	.10	.10

The average propensity to consume housing appears to have declined over the four years in question. If true, this finding implies a more rapid increase in nominal incomes than in nominal rents during this period. This inference is called in question, though, by the considerable discrepancies in the annual rates of change of the various income components over the period. Here is another instance of the liability to make mistakes in affordability calculations. This liability exists even in cases such as this one, where the samples have been carefully drawn and the surveys are of high quality. Given that this is so, it is better to rely on a few sound estimates, rather than proliferate crude ones.

Tables 3.2 and 3.3 provide results for all low-income dwellers in the Santa Ana and Sonsonate surveys, incorporating owners by imputing a rental value of their homes. This procedure yields much higher values of "a" for residents of colonias ilegales than for mesones dwellers.

Table 3.2
 THE AVERAGE PROPENSITY TO CONSUME HOUSING (SANTA ANT)*
 OWNERS AND RENTERS: 1980

	Whole <u>Sample**</u>	Meson <u>Renters</u>	Tugurio <u>Owners</u>	Colonia <u>Owners</u>	<u>Owners</u>
$a_1 = R/Y_{NH}$.20 (.33)	.17 (.22)	.11 (.09)	.43 (.69)	.39 (.64)
$a_2 = R/Y_{WT}$.12 (.24)	.09 (.06)	.07 (.06)	.30 (.61)	.27 (.57)
$a_3 = R/Y_{ST}$.12 (.16)	.09 (.06)	.07 (.06)	.35 (.77)	.24 (.34)
$a_4 = R/Y_{TT}$.11 (.11)	.09 (.05)	.07 (.05)	.24 (.23)	.22 (.22)

* For owners, an imputed rental value is used

** Weighted

Table 3.3
 THE AVERAGE PROPENSITY TO CONSUME HOUSING
 (SONSONATE)* OWNERS AND RENTERS: 1980

	<u>Whole Sample</u> **	<u>Meson</u> (Renters)	<u>Colonia</u> (Owners)
$a_1 = R/Y_{WH}$.18 (.18)	.15 (.15)	.34 (.27)
$a_2 = R/Y_{WT}$.13 (.12)	.12 (.11)	.22 (.14)
$a_3 = R/Y_{ST}$.13 (.12)	.12 (.11)	.22 (.14)
$a_4 = R/Y_{TT}$.11 (.08)	.10 (.07)	.18 (.10)

* For owners an imputed rental value is used.

** Weighted

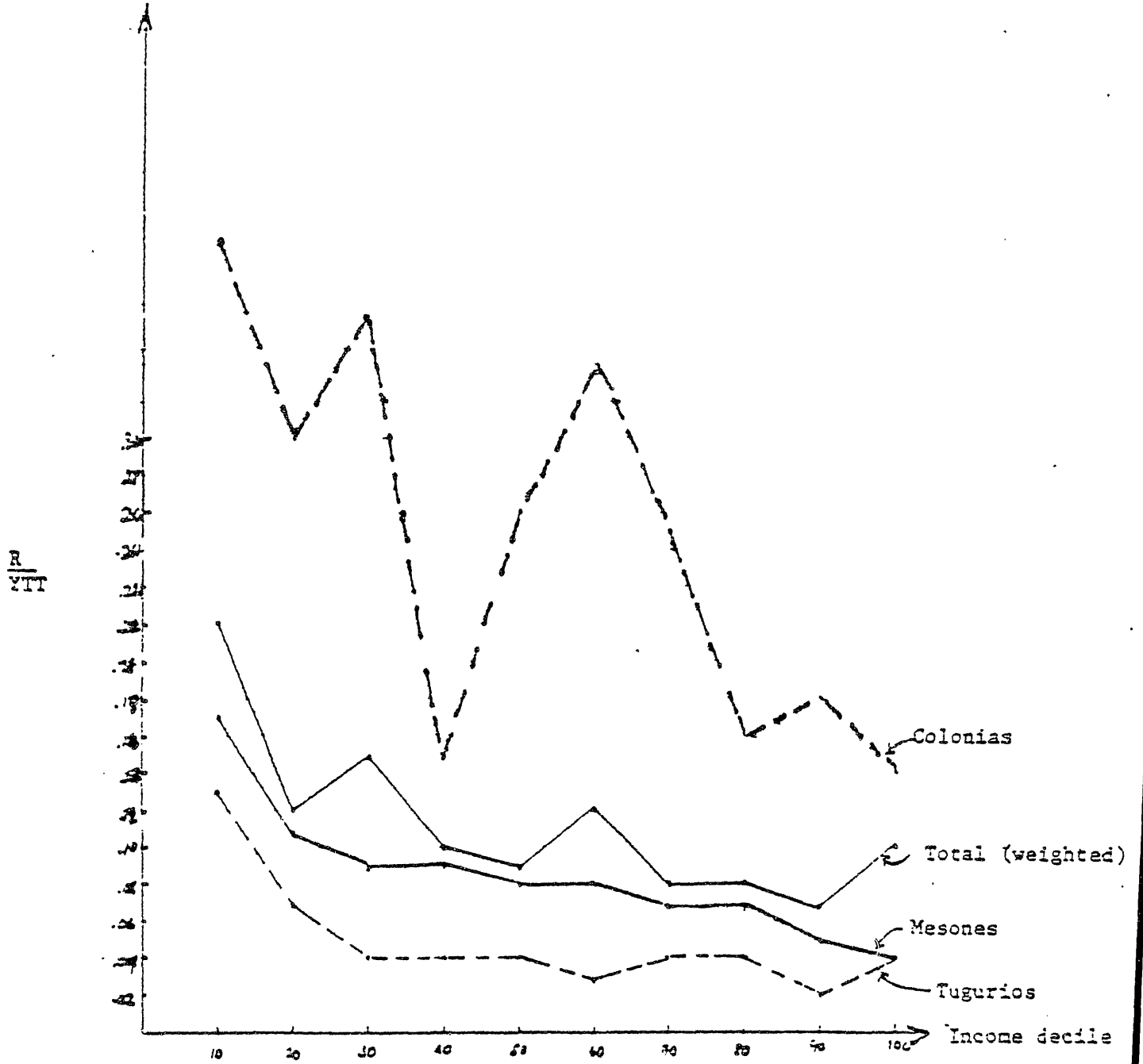
Those living in tugurios appear to behave more like renters. These results are illustrated in Figures 3.2 - 3.7 which reproduce respectively the scatter diagrams for Santa Ana (total, colonias, mesones, and tugurios,) and the plots of the income decile mean values of "a" for both cities by category.

The general result is that owners exhibit a greater average propensity to consume housing than do renters. However, further study will be needed before confident statements can be made concerning either the magnitude of or the reasons for this difference. Generally speaking, the difference is thought to result partly from definitional ambiguity, ^{1/} in part from irreducible data problems (the imputation of rental values), and in part from real behavioral differences. Future studies will be most concerned with this last order of difference.

It has typically, though not always, been assumed that the average propensity to consume varies systematically with income. Figures 6 and 7 generally confirm the downward trend hypothesized by Grimes five years ago. This is so despite certain shortcomings of the data: the relatively small size of the sample generates extreme values for a few cells, particularly for the stratifications by residential type, and as with all other studies of this type, this phenomenon is most marked--and the results correspondingly least certain--for the lowest 2 to 3 deciles. However, the following results appear

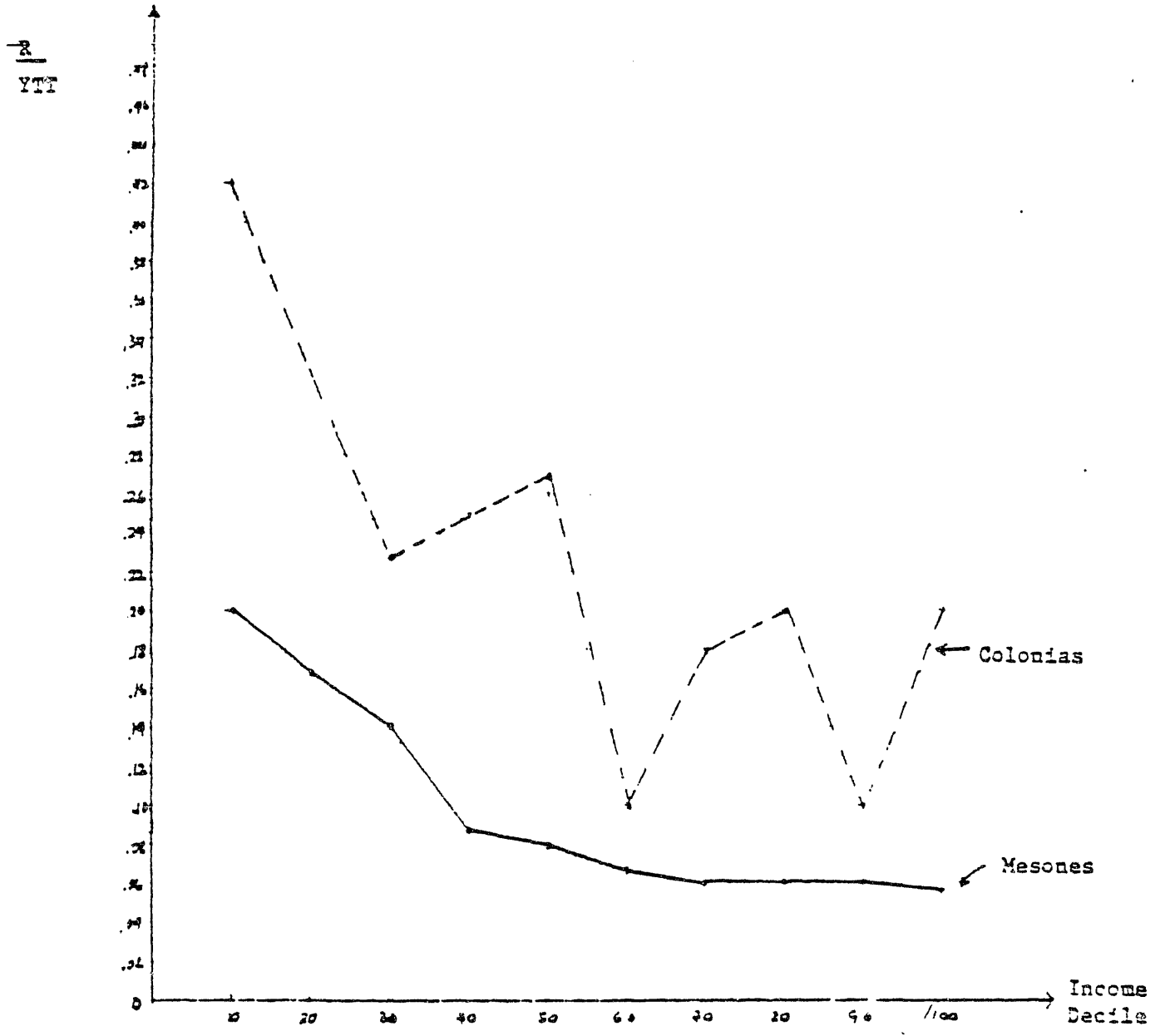
^{1/} This refers principally to the extent to which owners' housing outlays are for own housing consumption, as opposed to the creation and maintenance of rental or business accommodations.

Figure 3.6



SANTA ANA 1980

Figure 3.7



SONSONATE 1980

unambiguous: first, for both cities the total sample values decline, from .21-.22 for the bottom decile, quite abruptly through the first three deciles and then fluctuate in a narrow range between .10 and .08. Second, the stratified sub-samples adopt a similar pattern, such that those for colonias are highest, followed by owners, total sample, renters (mesones), and tugurios.

These trends appear clearly in the Figures. ^{1/} It should be noted though that they reflect only the particular stratification of data according to income, the observations having been grouped by tenure type. A rough stratification of the data according to housing quality alone reveals that the trend could be reversed:

	<u>Average Total Income</u>	<u>A₄</u>
Tugurio	¢ 320 per month	.06
Meson	¢ 470 per month	.07

^{1/} These findings also compare closely with estimates by Ingram (1980) and Renaud, et. al.(1978). [Refs.] Whilst these three treatments are not strictly comparable, because the other two studies had as their samples the entire urban populations of Bogota and Cali, Colombia, and of Korea, respectively, they have all found that the average propensity to consume housing declines with income.

We have found that this declining trend is most pronounced for the lowest three deciles of income, and is imperceptible for approximately the highest three deciles. The latter phenomenon is of little interest in the planning of housing projects for the urban poor. However, a more precise appreciation of the lowest decile's trend is of considerable importance and will be the subject of further study.

It is thus important to consider micro-level data which are not stratified--or, alternatively, where the implications of the stratifications chosen have been very carefully considered--in order to discover the true underlying relationships. 1/

The extensive literature on housing demand in developed countries has been reviewed by de Leeuw (19), and more recently by Mayo (1978). The general conclusion is that "for a wide range of analyses employing different data bases and methodologies, the permanent income elasticity of demand for housing is estimated to be well below one on average. Studies that have used data based on individual observations universally produce estimates below unity for both renters and owners." 2/

The previously cited papers by Ingram (1980) and Renaud, Lim, and Follain (1978) provide the first rigorous attempts to conduct housing demand estimation in developing countries. The latter analyzed data collected in 1976 in Korea for the regular Family Income and Expenditure Survey, and a Special Housing Survey. Ingram has used data collected in Bogota, Colombia in 1972, as part of the Bogota Phase II planning exercise, and a subsequent major research project conducted by the World Bank with Colombian assistance. Both studies are free from the aggregation and specification biases that had plagued many early studies in the United States. The approaches taken in this book are

1/ For a more complete treatment of this subject, see Ingram, op.cit.
2/ [Ref.]

generally very similar, though different in some specifics. They produce strikingly similar results, results which, though they deal with samples with widely varying incomes, are consistent both with one another and with comparable studies carried out using U.S. data.

These results are presented in Table 3.4. They indicate that the income elasticity of demand, as well as the average propensity to consume are quite stable across these countries. Although the results are not reproduced here, in the cited studies too, the average propensity to consume housing starts out high at low incomes, drops off rather sharply, as the minimal consumption of housing services varies little at low levels of income, and then appears to tend asymptotically to some lower value. According to the Korean data, this lower value is on the order of 8 to 12 percent, the relevant range is from 12 to 25 percent and the intercept may be inferred to be on the order of 30 percent or more. The Bogota data reveal that the highest income groups' average propensities to consume fall within the range of 12 to 16 percent, whereas the intercept may again be inferred to be above 30 percent.

The last stage of our analysis to date has been to trace the determinants of housing consumption in order to try to understand what lies behind the observed trends. The analysis is similar to that of the growing body of work on housing demand. However, there are a few important differences. One is that, because of the nature of our data base, we have been able to obtain a better measure of permanent income, which is estimated as a simple average over a three year period. A

Table 3.4

	Follain Lim and <u>1/</u> Renaud	Ingram <u>2/</u>		
		Log-log	Linear	
Income Elasticity of Demand	.6	.78	.35 to .73	Rises with income Rises with city size
Owners	.6			
Renters	.4			
Price Elasticity of Demand	-.2 to -.3 <u>3/</u>	-.86	-.63 to 0.85	Rises with income
Elasticity of Housing Consumption/Expenditure with Respect to:				
Family Size	.13 to .19 <u>4/</u>	.13	.12 to .28	Falls with income
Age of Head	n/a	.14	.07 to .23	Falls with income
Sex of Head	n/a	-.16	-.07 to -.16	Falls with income
Mean Income (\$) Year	\$1872 1972			

1/ Permanent income (consumption) equations only; national and/or sub-groups; log-log

2/ Household income

3/ -.60 to -.67 for cities other than Seoul

4/ Higher for current income and renters

second is that we postulate that different forms of income contribute in varying proportions to the decision to consume housing. We wished to investigate the hypothesis that the wage income of the household head (Y_{WH}) was the primary determinant of consumption behavior. So, in contrast to other studies, we have included other measures of income in the sample. ^{1/}

The housing equation estimated is of the linear form, and for the "i"th household is written: $E_i = Z_i b + u_i$ where $E_i = P_H H_i$ = housing expenditures in terms of actual or imputed rent; Z_i is a row vector of explanatory variables, $Z_i = (Y_i X_i)$, where Y_i is a vector of measures of income and X_i is a vector of demographic variables; b is a column vector of coefficients; and u_i is a random error term. The results of the estimation are presented in Tables 3.5 and 3.6, the former using current income and the latter permanent income measures for Santa Ana.

The first observation is that each of the "other" income variables, representing sources other than the wage income of the household head, contributes significantly to explaining the variability of housing consumption. According to the current income measures in Table 3.5, the marginal propensity to consume housing out of "other" income exceeds that for all other income variables. This finding is consistent with observations for the Philippines and Senegal that house construction for a majority of families is financed by savings and

^{1/} Notably: wage income of other household members (Y_{WF}), other income of the household (Y_{OF}), and total household income (Y_{TT}).

transfers from the extended family network, rather than out of current wage income. These results are largely confirmed when permanent income measures are used in Table 3.6.

Further revealing information is contained in Table 3-7 which presents estimates of the income elasticities of housing demand for the various sub-groups at their respective mean incomes and rents. These results support the conclusion of the previously cited studies that demand for housing is relatively income inelastic. This is true for renters, colonia owners, and tugurio owners in Santa Ana, across all four categories of income considered. The elasticity of demand with respect to "permanent" wage income of other household members and with respect to other "permanent" non-wage income is stable across income groups. The elasticity of demand with respect to the three-year average of wage income of the household head is considerably higher for the renters of Santa Ana mesones than for dwellers of the colonias. A similar result is apparent with regard to total permanent incomes of the household. This is consistent with Ingram's results for Bogota. The main results of this analysis are that, even when controlling for demographic variables, if prices are constant, the income elasticity of demand is significantly less than one (in the range of .5 to .8) and the average propensity to consume falls as income rises, which is consistent with our earlier results.

Household size does not seem to contribute significantly to housing demand. However, the sign of the coefficients reflects Ingram's finding that increased family size leads to increased demand among

Table 3.5

HOUSING CONSUMPTION IN
SANTA ANA 1980--CURRENT INCOME

	OWNERS		RENTERS		ALL	
CONSTANT	131.933** (2.846)	122.130** (2.894)	-40.780 (-2.235)	-20.828 ⁺ (-1.036)	-15.612 (-.7110)	2.240 (.1050)
TYT		.0912** (3.451)		.1875** (9.142)		.1570** (8.900)
YWH	.0732* (1.826)		.2420 (11.864)		.1932 (8.541)	
YWF	.0972** (2.593)		.0720** (2.659)		.1039** (4.121)	
YOF	.1076** (2.485)		.1145** (2.441)		.1609** (4.834)	
AGEH	-.0318 (-.0430)	.1494 (.2240)	.2694 (.9631)	-.1619 (-.5386)	.4785 (1.375)	.2159 (.6458)
SEXH	-19.432 (-.836)	-25.383 (1.249)	-23.206** (-2.334)	-.6786 (-.0615)	-22.832 (-1.890)	-9.590 (.8496)
HHSIZE	-7.264* (-1.551)	-6.854* (-1.610)	.3589 (.4613)	-4.477** (1.979)	-.6579 (-.2544)	-3.660 (-1.5213)
R ²	.16	.16	.54	.40	.33	.30
N	72	72	136	136	208	208

* significant at 95%

** significant at 90%

● coefficient exceeds standard error

Table 3.6

HOUSING CONSUMPTION IN
SANTA ANA 1980--PERMANENT INCOME

	OWNERS		RENTERS		ALL	
CONSTANT	113.462** (2.343)	100.524** (2.204)	16.965* (1.730)	9.990* (1.043)	.4133 (.0161)	-6.364 (-.263)
PY		.0895** (2.795)		.0295** (2.566)		.119** (5.617)
PYWH	.0681 (1.324)		.0197+ (1.193)		.1059** (3.370)	
PYWF	.1113** (2.439)		.0111+ (.950)		.1142** (3.963)	
PYOF	.1012** (2.019)		.1048** (5.618)		.1522** (4.263)	
AGEH	-.0802+ (-1.966)	.1620 (.234)	-.0743 (-.538)	.1304 (.964)	.279 (.704)	.4119+ (1.121)
SEXH	-7.683 (-2.98)	-18.803 (-.851)	7.021+ (1.524)	5.471+ (1.219)	-6.437 (-.474)	-8.906 (-.738)
HHSIZE	-4.713+ (-1.012)	-3.576 (-.829)	.8850+ (1.003)	-.1317 (-.1424)	.1358 (.0526)	-.140 (-.058)
R ²	.13	.12	.29	.10	.22	.21
N	63	63	93	93	156	156

* significant at 95%

** significant at 90%

+ coefficient exceeds standard error

Table 3.7

INCOME ELASTICITIES OF DEMAND FOR HOUSING
(EVALUATED AT THE MEAN)

	Santa Ana		Bogota		Cali		Korea	
	Renters	Owners	Renters	Owners	Renters	Owners	Renters	Owners
<u>Permanent Income</u>								
Wage of Head	.1599	.1589						
Wage of Others	.0600	.2316						
Other Income	.1285	.0647						
Total Income	.4338	.4530	.78*	.75*	.47*	.73*	.42	.62

* Current income estimates.

renters, but not among owners. This would be especially true for meson units which tend to be one-room dwellings. The only other significant demographic variable is sex of household head. Our estimated coefficients indicate that female-headed households tend to demand more housing in mesones and colonias and are thus consistent with Ingram's findings for Colombia.

Principal Conclusions and Remaining Questions

World Bank evaluation studies and comparable recent research have found that "a" basically declines as income rises, because housing is a necessity. In addition, "a" varies with other factors, including household characteristics, city size, and nature of tenure. Owners tend to consume considerably more housing than renters, and as income rises, the ratio of owners' to renters' consumption rises. Even if we adjust the calculation for owners to ensure strict comparability by adding rent implicitly to their incomes so that, for owners, $a = \frac{R}{Y + R}$, only part of the discrepancy disappears. This furnishes evidence that homeowners own their homes not just for annual housing services provided, but for some combination of other reasons, including use of (part of) the premises to earn income, tax advantages, expected capital gains, etc.

In researching affordability, the measure of income chosen is crucial. The disaggregated results of World Bank experimentation with urban housing show that housing expenditures are least responsive to the measure of income originally favored--wages of household head--and much more responsive to other income components. As we have emphasized, most

important for project design purposes is total household income, including "unearned" as well as earned income sources. Total household income comprises a significant indicator of family purchasing power, providing a more encompassing definition that reduces the risk of screening out people who should and can be helped by housing programs because they can actually afford them. Also contrary to original expectations, total household income is less subject to variation than is wage income taken alone. This measure of income must still be used with care, and attention paid to the distinction between transfers that are stable components of income and those that are not. World Bank data, particularly those from the Philippines, enable us to provide reliable guidance on this subject.

The elasticities estimated for the various cited studies of housing consumption are strikingly similar, falling within a narrow range and varying systematically as between renters and owners. This does not imply identical values of "a" between countries and cities. In fact, the values of "a" appear to vary considerably between countries and also to vary with city size. Another way to put the matter is to say that, while the curvature (elasticity) of a/Y curves appears to be quite stable across various situations, the height (as measured by the intercept) seems to be more variable. It is perhaps more accurate to say that we have little reliable information yet about the intercept or, for that matter, about what the a/Y relationship looks like in the lowest 2 to 3 income deciles. The variance appears to be much higher in this range than elsewhere, and we do not yet have a clear idea of the

degree to which this can be attributed to instability of income, as opposed to other household characteristics, or of how the effects of these variables relate, for example, to those of city size and climate.

The next phase of work will concentrate in this area, using data from project populations as well as controls, and from additional countries, in order to increase the accuracy of measurements. While we are seeking more reliable, eventually cheaper measures, it will generally be wiser to transfer good parameter estimates from comparable environments than to attempt to estimate parameters using bad data from the site, even though much remains to be learned about defining "comparable" environments and about "transferability". It is worth noting that in seeking measures that are cheaper and easier to apply, it matters very much whether we are still refining our understanding of how "a" varies with "Y" or using improved measures of "a" for project design purposes. When working on the former, we continue to utilize detailed information from carefully drawn samples. Once satisfied with the measures of "a", we can apply them to relatively crude estimates of income distributions to set affordable limits to project costs.

Evaluation research to date suggests that the World Bank's major urban housing projects have been affordable to their target populations. This has been so even though actual costs have turned out considerably higher than originally estimated. There are two sets of reasons behind higher costs. First, materials have risen in price at unforeseen rates, and the impacts of this factor and other inflationary

forces have been intensified by substantial delays in project execution (averaging roughly 18 months). Second, private costs (what families have voluntarily invested in their plots and dwellings) have greatly exceeded initial estimates. This trend has resulted partly from inflation and partly from the fact that families have generally disposed of more resources for house-building than had been anticipated. Notwithstanding these factors, it is heartening to find that projects have been affordable even to families in the bottom 20 percent of the income distribution. This is not the same as saying that distributional impacts of the projects have been optimal. The latter subject is treated in a later chapter.

The encouraging findings on affordability are supported by indirect indicators related to possible affordability problems. One such indicator would be uncommonly high turnover of project populations. This is not an unambiguous indicator, as people would be moving out at high rates for reasons unrelated to affordability. For example, they could be realizing substantial capital gains. On the other hand, they could be staying on in spite of affordability problems, in which case the effects would likely show up eventually in one or more of the following indicators. But in none of the projects studied has turnover among project populations been higher than that among controls. In El Salvador, the turnover rate has been much lower, allowing for some early departures amongst families that found participation in the mandatory mutual help program onerous.

Another possible indicator would be an unexpectedly slow rate of house consolidation. This phenomenon has been observed in the Senegal project, and initially affordability was involved. Government spokesmen raised people's expectations about what would be done for them, whilst the project initially had no loan component. When one was introduced, it was inadequate. Even since these problems were resolved, project progress has been slow, for reasons having to do with more general design failures to be discussed later. Construction in all other projects has proceeded more rapidly than expected.

Still another possibility is that families might attempt to adapt to affordability problems by altering income and/or expenditure patterns. There has been particular concern in the early phases of the projects that participants might be forced to reduce consumption of other necessities, such as food or health services, to meet housing payments. In fact, there have been adjustments on the income side: participating families have a high propensity to expand their houses and rent accommodation, and there have been positive effects on labor force participation and other economic activities. These, however, can hardly be considered indicators of affordability problems so long as the families are successful. On the other hand, there have been no indicators that investments in housing have led to reduced consumption of food, medicine, or other necessities.

A final indicator would be a relatively high rate of default, which is linked to the first indicator, as, presumably, households whose payment lags persist will eventually be evicted. Here the evidence is

more mixed. In El Salvador, default rates are virtually nil. In the Philippines, collections have yet to start. In the two African countries, however, default rates on plot payments (and in Zambia on loan repayments) have been distressingly high. In the case of Senegal, it is difficult to extricate this phenomenon from the complex of problems that have plagued the project; however, it does not appear to indicate primarily of affordability problems. In Zambia, it was initially thought that affordability might be the key factor, particularly given the economic reversals faced by the country during the period of project execution. Further research however, has shown this not to be the case. Defaulting has not been related to income, so much as to project design and execution. This issue too will be treated later.

The implications of affordability for project design are developed in detail in the following chapter, but it is worth reiterating here the major implications of affordability analysis. In essence, we want to design housing projects that will "reach" as far down as possible into the income distribution. This makes admissions criteria necessary, but this is not a simple matter. It is possible to err in several ways: by pushing the lower income cut-off too low, we may admit households that will fail to make adequate payments; if the lower income cut-off is too high, poor households that might succeed may be rejected. Towards the upper end of the target distribution, it may be desirable to encourage some participation at higher prices to produce

favorable cross-subsidy and market interactions. At both ends of the spectrum, affordability criteria and judgments remain to be continually improved.

Chapter 4:

MEASURING PROJECT IMPACTS

Experience has brought the greatest changes in evaluation procedure in the area of measuring impacts. The Bank's evaluation program was initially sold principally on the grounds that it would provide quantitative evidence of benefits deriving (initially) from sites and services interventions. Very little was known in the early '70s about such matters. But there was an unfortunate failure to be decisive about the benefits or impacts to be measured early on.

The pilot program was developed by consensus, and this consensus was achieved, and at first sustained, by attempting to measure virtually everything that interested anyone. The program and research instruments were designed to measure (or at least permit subsequent evaluation of) all the major socio-economic impacts the shelter projects were expected to have. An illustrative list includes: stimulation of housing construction and savings; conversion of sweat equity into housing construction; generation of building skills that would lead to enhanced future income for beneficiaries; improvements to housing and neighborhood environments; linked improvements to urban management and urban form, etc.

These manifold tasks were adopted in the hope that insights could be gained in all directions. In fact, they could be - that was not the problem. The problem was that because certain phenomena take much longer to develop than others and certain relationships among them are much more complex than others, the cost differentials for improving

our understanding of the many questions vary enormously. This was anticipated, of course, and discussed at length during planning stages. However, because of the pervasive ignorance about the whole field at the time, there was little basis for confident decision or agreement upon the most fruitful areas of research opportunity. Lacking answers, and impatient to get underway, individuals and institutions forged ahead. In retrospect, this was a serious error. It led to an undesirable diffusion of effort, to survey instruments that were considerably longer than they might have been, and eventually to a problem of data "overhanging."

As the program progressed and attempts were made to launch rigorous economic analyses into subject areas such as employment and health, the necessary set of priorities regarding timing began to emerge inexorably. In the field of health, for example, the relationships suspected to be involved were found to be so numerous and complex that any research effort worthy of the name would be very costly and very risky. An accumulation of experiences of this sort during the program's first year-and-a-half led to a rethinking and reformulation of our approach, guided by the literature. ^{1/}

Phenomena to be considered would now be divided into four categories: inputs, outputs, effects, and impacts. Inputs are the factors of production: labor, capital, and management. Outputs should

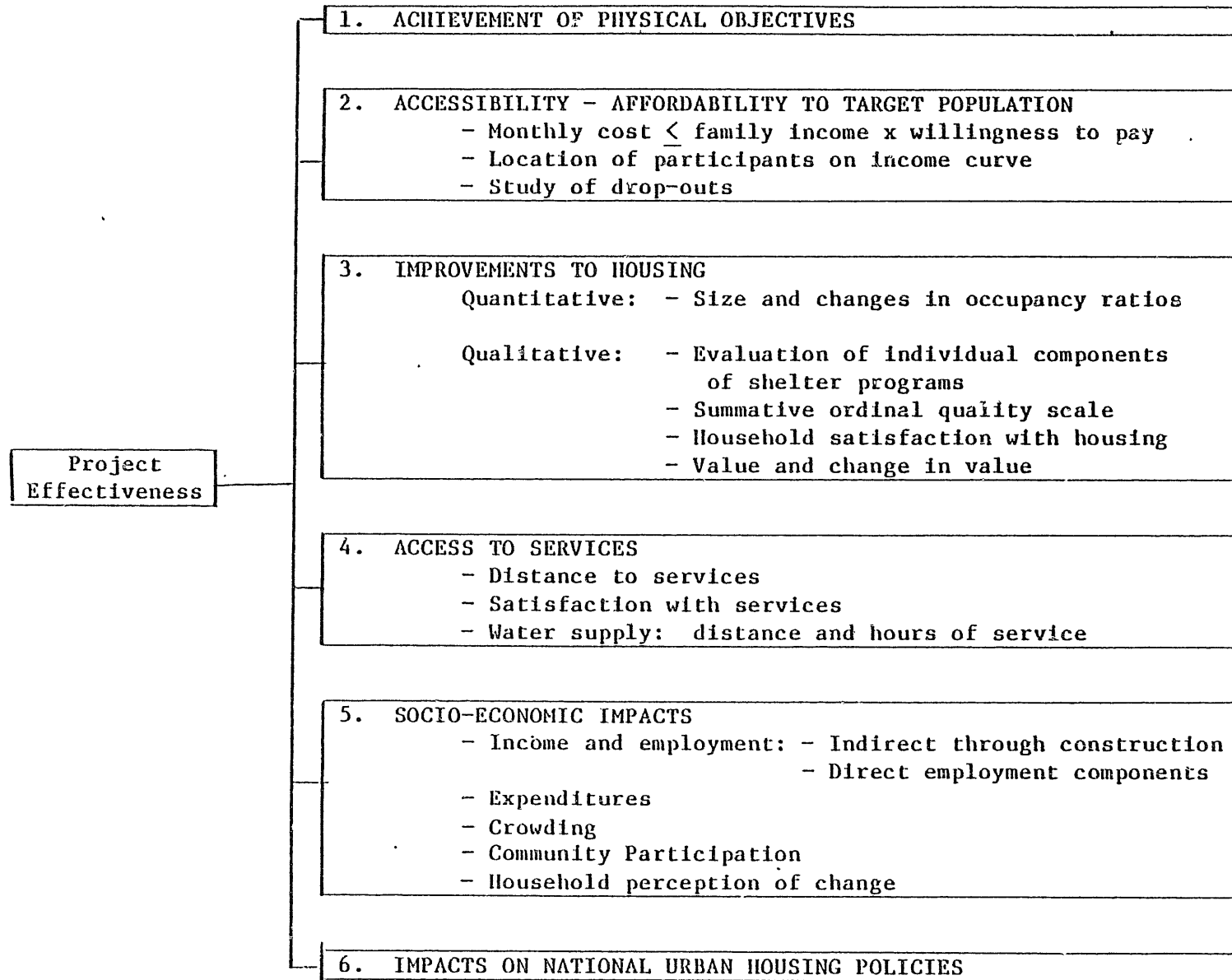
^{1/} Remi Clignet and Michael Bamberger, in particular, modified methodological features of the evaluation program in accordance with the directions of leading evaluation practitioners such as Donald Campbell, Thomas Cook and their associates. See chapters 2 and 9 of this volume.

be recorded in two ways: as the physical achievements of projects, and as performance standards. The distinction between effects and impacts is less clear cut, but it is useful to try to maintain it in the interest of avoiding the types of problems described above. "Effects" refer to the more-or-less direct and immediate results of a project's execution, e.g., actual increments to housing, service access, and income, for sites and services projects. "Impacts" refer to longer term, less direct results of a project's existence, such as sustained improvements in community welfare (income, health, education, etc.) resulting from a sites and services investment. One might consider making the distinction between effects and impacts on the basis of time alone. Alternatively, it can be viewed as a question of assessing the difficulty and cost of obtaining results which meet external validity tests.

This chapter discusses a range of project results under the rubric of "impacts," though it is arguable that many of the findings treated could just as well be categorized as "effects."

Although the four projects in the evaluation program emphasize different features, all have been centered on providing or augmenting affordable, appropriately serviced housing built according to the concept of progressive development. The evaluations have so far gauged impacts of the projects in six areas: (1) achievement of physical objectives; (2) accessibility to target populations; (3) improvements to housing; (4) access to services; (5) employment and income generation;

Figure 1: CRITERIA FOR ASSESSING PROJECT IMPACTS



and (6) broader impacts on national urban housing policies and on urban areas. (See Figure 4.1)

Achievement of Physical Objectives

The projects have largely attained their stated physical objectives. In El Salvador, over the past five years, despite land acquisition problems, 6594 houses have been completed or are in progress. This constitutes roughly 30 percent of the formal housing built during this period. The basic units provided by the FSDVM were erected at an average cost of 3383 colones, or roughly US \$1350, whereas other public housing programs in El Salvador have been devoted to building substantially more expensive housing for a considerably wealthier segment of the urban population. (In the period 1973-78, for example, the Salvadorean Foundation for Housing (FSV) constructed 5000 units costing between US \$4000-6000 expressly for those workers covered by the country's social security system, and hence available only to the wealthiest 30 percent of the population participating in that system.) Further, the housing quality achieved in the projects has been as good or better than other market alternatives, with the sole exception of floor quality. Compared to other informal market alternatives, project housing has been as good or better (than either tugurios or colonias ilegales). Desired levels of water supply and drainage have nearly been completed, although a lack of government agency support has caused shortfalls in the construction of the complete package to include schools, clinics, and community centers. The FSDVM has also built most of the infrastructure it was responsible for, including a contingency

well, storm drains, and footpaths, though lack of demand led to the decision against completing a proposed market.

In Zambia, half the population of Lusak has been affected by the first project, and the production of low-cost housing has been increased by 50 percent. As a result, the living environment of nearly 31,000 houses has been improved by upgrading basic services. Some 17,000 families have received new installations of piped water, roads, and security lighting. The housing stock itself has been increased by 11,300 through both sites and services and "overspill" provisions. (Overspill areas are upgraded zones adjacent to improved squatter areas; they include households displaced by the project as well as provisions for absorption of future urban growth.) A sizeable proportion of the schools, community centers, and markets planned for the Lusaka area has been completed. Delays due to inflation and cost escalation have hindered the completion of those services and the provision of health centers. Additional funds have been approved for completion of these components.

The projects yet to be completed have also witnessed progress towards realization of physical objectives. In the Philippines, 97.5 percent of sampled households across a relatively wide range of incomes have undertaken some kind of housing improvement, with 12.5 percent building completely new structures. Although reblocking has reduced the average lot size in the Tondo project area from 65.2 to 53.9 square meters, the range of lot sizes has been made considerably more equitable. Concurrently, the housing size range has shifted from a

minimum 13.7 square meters and a maximum of 400 square meters, to one concentrated between 32.9 and 88.4 square meters. Rental space in reblocked areas has also increased. An average increment of 10.5 square meters has been added to sublet space; the number of families offering rented rooms has itself increased by 10 percent. This finding illustrates the importance of timing in the evaluation of such programs. In the early stages of reblocking, the number of renters dropped substantially. ^{1/}

Despite serious problems, including excessive design standards, delays in provision of infrastructure, inflation, and its location a considerable distance from the city center, the Dakar project in Senegal has tripled the population living in its site within the past year. With acceptance of the core construction method by participant families (which includes construction loans and the use of project-approved contractors), households have begun to move into two-room homes on the site and to plan additions to those homes in the later phases of construction. Provision of additional water supply, lighting and security services should lead to an accelerated rate of project site occupation.

Accessibility to Target Populations

Measures of accessibility have indicated that the projects have been able to incorporate low-income populations, despite "encroachment" by middle income groups. In El Salvador, at least 85

^{1/} House Consolidation Study, Tondo Foreshore Dagat-Dagatan Developments Project Report Series 80-2, Research and Analysis Division, National Housing Authority, Philippines, 1980, p.49.

percent of sites and services participants come from the lower 65 percent of the urban population, with the majority representing the third, fourth, and fifth income deciles. Although initial project stages witnessed a substantial dropout rate (mainly due to mandatory participation in the mutual help program), once occupied, the project sites have shown very little turnover, markedly less than in either the non-project control areas or the market as a whole. The evaluation has also shown, as previously cited, that families who do leave are not the poorest, but come from all income strata, and they do not appear to be replaced by families with higher incomes.

The Zambia project has fared even better in reaching low-income targets. Sixty percent of the population of George, a typical upgraded area, come from the poorest 20 percent of the urban population. Fifty eight percent of the population in Lilanda, a "normal" sites and services project, are representative of the poorest 30 percent of urban families. The Lusaka project does not appear to have encouraged rural migration to the city, contrary to fears expressed at the project's initiation.

As noted earlier, the Tondo project has seen few cases of legitimate turnover, although an interim report has raised the question whether a significant proportion of families will be able to afford reblocking. Some illegal selling of rights to tagged (censured) structures in the project areas has been observed.

In the ongoing Senegalese project, increases in costs and the fact that living standards haven't risen have spontaneously discouraged

applications for participation from low-income households, according to the Bureau of Evaluation. While in Lusaka, turnover amongst families moving to overspill areas, though not great, is skewed, with younger, smaller and economically weaker families (i.e., "natural" renters) being overrepresented. Project findings thus point to the need to continue scrutinizing selection procedures.

The foregoing observations indicate that the projects have been broadly successful in their targetting. It remains apparent that considerable improvement is still possible. In order to better understand the characteristics of those households that populate the informal housing sector, a study was made of the income profiles of such communities. Using data collected in three countries--El Salvador, the Philippines, and Zambia--it was possible to determine the relative position of project beneficiaries in their respective income distributions. The major finding is that in both sites and services and upgrading projects, the target populations span a wide relative income range and tend to be more representative of median income than of the poorest urban households.

Table 4.1 presents the available evidence. In each instance, data on the incomes of project households have been matched with national urban income distribution statistics. The project income profiles are meant to reflect the situation prevailing before project implementation. Our attempts at an ex-ante portrayal thus try to abstract from any income-augmenting impact the projects themselves may have had. The table presents the percentage of beneficiary households

that falls within specific quintiles of national urban income distributions. Specific magnitudes must be treated cautiously, as both measurement errors and methodological limitations constrain the accuracy of any given number.

Line "c" of the table presents the percentage of beneficiary households who lie above an arbitrarily selected "poverty line" drawn at the 40th percentile. The projects reported on tend to have close to half their participants belonging to this upper 60 percent. This finding leads to the conclusion that median income households are representative of beneficiary populations. The only exception to this conclusion is the Zambian upgrading project which clearly serves the poorer strata of Lusaka's population.

The three things most surprising about these findings are: the similarity of income profiles between sites and services and upgrading projects, the extent of penetration into the bottom 20 percent of the income distribution, and the substantial percentages in the top 40 percent of the distribution. Line "a" indicates that, relative to sites and services, upgrading projects reach a larger percentage of the poorest urban residents. Project designers anticipated this outcome, but the extent of the differences has fallen short of most expectations. At the opposite end of the spectrum, according to line "e", all projects tend to benefit a sizeable number, 19 to 30 percent, of families with incomes above the 60th percentile. For sites and services projects, this "linkage" of some project benefits to a less needy group suggests that more attention should be paid to upper

Table 4.1

URBAN SHELTER PROGRAM
(In Percent)

National Urban Income Percentile	SITES AND SERVICES				UPGRADING	
	El Salvador		Zambia		Philippines (1979)	Zambia (1976)
	Sonsonate (1977)	Santa Ana (1976)	Lilanda (1980)	Matero (1978)		
(a) 0 - 20	6	11	28	18	27	38
(b) 21 - 40	38	32	26	38	24	22
(c) Upper 60	56	57	46	44	49	40
(d) 41 - 60	37	38	16	14	23	17
(e) > 60	19	19	30	30	26	23
TOTAL	100	100	100	100	100	100

boundary affordability criteria. For upgrading projects, the presence of a large upper income tail illustrates the income heterogeneity of squatter community residents. (Note that in addition to income heterogeneity, a squatter community like Tondo also contains a wide variance in housing quality.) This finding provides further evidence that informal housing is not synonymous with low-income housing, and that inadequate shelter may not be ascribable solely to the constraints imposed by absolute poverty. (In the case of Tondo, 26 percent of the beneficiary families reported total household incomes close to or in excess of US \$2000 per year.) The "linkage" of project benefits in this circumstance must be generally accepted as a consequence of in situ shelter development.

These empirical results are intended to reveal the distributional aspects of project benefits. The findings are analogous to what has been learned about employment in the informal sector, namely, that failure to belong to the formal sector is not a reliable indicator of poverty level. The results further recommend that we distinguish between housing objectives and poverty objectives when justifying particular interventions into housing markets. If poverty alleviation is a primary goal, it may be that the types of urban shelter projects we have experimented with are not the most effective mechanisms for satisfying such objectives. If improving efficiency in the housing market is a primary goal, then existing intervention schemes can be justified by placing less emphasis on their distributional outcomes. The point is that housing and poverty objectives may not be attainable

with the same interventions, and that these goals should be addressed individually during project design and implementation.

The relative success of a project depends to some extent upon the emphasis of the evaluation study. If the focus is on low-income penetration, the Zambia project is very successful compared to the El Salvador project. On the other hand, if high income exclusion is the criterion, the El Salvador project scores better. Keeping in mind that the El Salvador approach (sites and services) can be more finely honed with respect to its distributional objectives, whereas the Zambia one (upgrading) scarcely can be, the foregoing suggests a new way of looking at sites and services vis a vis upgrading tactics. Previously, squatter area upgrading has been seen as a vastly superior approach to "reaching the poor." Operations management was for some time asking evaluators whether the Bank should not concentrate exclusively on this approach. The available information suggests that the answer depends upon how objectives are defined. If the goal is to achieve maximum penetration into the lower income deciles, then area upgrading is the superior approach, though it permits substantial "leakage." If the goal is to concentrate benefits within a narrowing range of low (but not the lowest) incomes, then sites and services schemes look more promising. The determinants should be an explicit weighting of benefits to various deciles and success in improving sites and services targetting.

Improvements to Housing

Estimations of housing quality have been undertaken principally in El Salvador and the Philippines. In both countries, the

evaluations have attempted to employ measures which incorporate families' assessments of dwelling quality with non-subjective means, such as visual observation of building materials, housing's monetary value, and professional appraisers' value assessments.

In El Salvador, a scale was used combining three-point ratings of quality of walls, roof, floor, sanitary facilities, and water supply, with values converted to percentages to simplify interpretation. (Such scales are suggested as a tentative but useful means of summarizing quality tabulations; resulting figures represent only a relative order of magnitude, not a genuine interval scale.) Three types of informal housing were ranked on this scale at project initiation in 1976, and again in 1980. Table 4.2 shows that in 1976 very little difference in total housing quality was found between mesones and colonias ilegales (67 and 71 percent, respectively), whereas tugurios ranked lower on all indicators (22 percent). By 1980, housing quality of the colonias ilegales and tugurios had registered improvement, whereas no such change in the mesones was indicated.

Given that nearly 83 percent of the project participants in Santa Ana formerly resided in mesones, the 1980 results comparing participants with residents continuing to live in mesones, colonias ilegales, and tugurios provide a good indication of the housing quality participants would have enjoyed if they had continued to live in mesones. They also report the changes in the physical environments of the participants. As Table 4.2 shows, a move to the project, during the period would have yielded a 23 point increase measured against the

weighted average of the other options (29 points against mesones). A family that moved from a meson to a colonia ilegal would have enjoyed an improvement in housing quality of 11 points, rather than having housing quality stay roughly the same in the meson. The only measure of housing quality involving a decline (and that a slight one) for participant moving to the project was the floor. Projects have so far eschewed providing cement floors in order to keep costs down, and because families appear to accord flooring a lesser priority than other housing features. As for satisfaction with lot size, living area, housing materials and overall quality of construction, the project ranked higher than mesones, colonias, or tugurios.

Attempts were also made to estimate changes in housing value. These changes cannot be measured directly at market values, given project stipulations (in El Salvador for instance) that houses not be sold for a minimum of five years after completion. As shown in Table 4.3, five possible indicators of housing value in 1979 were used as proxies: 1) construction cost, 2) owner's calculation of scale price, 3) owner's estimate of rental value, 4) comparison of values and costs, and 5) comparison of rent (multiplied by 100) with value. (In a well functioning market, it is often assumed that monthly rent equals one percent of a house's value.)

The table reveals that the mean cost of colonia houses is more than twice as great as that of project housing, but with little difference between the two in the lowest quartile. The finding provides another indication that the colonia covers a much broader spectrum of

Table 4.2

QUALITY SCORES IN PERCENTAGE FORM FOR LOW-INCOME
HOUSING IN SANTA ANA - 1976 AND 1980

	<u>Roof</u>	<u>Walls</u>	<u>Floor</u>	<u>Water</u>	<u>Sanitation</u>	<u>Light</u>	<u>Average</u>
<u>1976</u>							
Meson	98	51	76	50	36	92	67
Colonia	88	79	57	66	50	87	71
Tugurio	28	17	1	32	9	42	22
<u>1980</u>							
Meson	99	55	53	47	47	95	66
Colonia	93	79	41	84	76	96	78
Tugurio	48	37	2	49	34	40	35
Project	100	99	48	100	100	100	91
<u>Change 1976-1980</u>							
Meson	+1	+4	-23	-3	+11	+3	-1
Colonia	+5	0	-16	+22	+24	+9	+7
Tugurio	+20	+20	+1	+17	+25	-2	+13
Project	+1	+46	-32	+51	+55	+6	+21

Note: The percentage scores are based on the frequency distribution given in Table 4.1 of the main report on El Salvador. These were transformed into percentages where 100 percent indicates all families had good quality and where 0 percent indicates all families had bad quality. The scores are intended as a simple way to summarize the results; the figures only represent a relative order of magnitude rather than a genuine interval scale.

Source: Bamberger, Sae-Hau and Gonzalez-Polio, op. cit., forthcoming as Bank Staff Working Paper in 1982.

housing. Using owners' estimates of value, the value/cost ratio is considerably higher in projects than in colonias, suggesting a higher rate of return on investment in the project. A similar conclusion derives from the comparison of rents and value. In the project, rent is equal to 0.77 percent of value, whereas the ratio is 0.66 for colonias, again suggesting that a higher proportion of value is recouped in the project. It is interesting to note that in tugurios rent is 2.8 percent of value, again indicating that people are prepared to pay for the favorable location, even though uncertainty of tenure makes the expected sales price relatively low.

In estimating changes in value, a comparison was made for participants between the rent they paid in 1976 and imputed rental value of the project house in 1980. On the assumption that rent is proportional to value, the changes measured can provide an indication of the changes in a house's value. The mean value of imputed rent in 1980 was 3.41 times that for 1976 (adjusted for inflation). In contrast, rent increased only 1.45 times for families renting in colonias ilegales and 1.22 for families in mesones. The methodology used for these estimations is not very precise, but the figures do suggest that a move to the project was associated with a large increase in the value of housing.

Although shelter in various types of informal housing can be cheaper than in FSDVM projects, cost-benefit analysis investigating internal rates of return has demonstrated that the Fundación's projects

Table 4.3

COMPARISON OF COST, ESTIMATED VALUE AND RENT FOR PROJECT AND INFORMAL HOUSING
SANTA ANA 1979.

	Estimated Cost Value (1)						Rent (2)	Value/ Cost	Rent x 100/ Value
	\bar{X}	Interquartile Range		\bar{X}	Interquartile Range				
<u>Owners</u>									
FSDVM	7,639	5,718	8,743	10,800			83.3	1.41	0.77
Colonia	19,811	5,690	20,491	20,728	1,200	25,000	136.8	1.05	0.66
Tugurio	617	276	800	616	350	800	17.8	.99	2.88
<u>Renters</u>									
Meson				3,317			33.2		
Colonia				5,122			51.2		

Note: (1) For Owners, Value = Estimated Sale Price
For Renters, Value = 100 x Rent

(2) For Owners this is their estimate of Rental Value

yield higher rates of return than any other type of formal or informal housing available to low and middle-income groups in El Salvador. For a family living in a meson, from which most participants come, the move to the project in fact triples the rate of return. Project housing also ranks most favorably when the options are scaled according to net present value, and net present value as a fraction of total costs. It appears that a family can buy more housing services (that is, receive more benefits) for a given sum from the FSDVM than from any other available shelter market options. (See Table 4.4.)

In the Philippines, various techniques have been used to estimate housing value, and hedonic pricing techniques have been used to determine which characteristics of squatter dwellings contribute most to housing value, as well as to obtain price-consistent measures of change in housing quality. The former have demonstrated that squatter dwellings are houses with considerable value, the average appraised value being about double average annual income, which is about the same ratio as for formal markets in the U.S. and elsewhere.

Hedonic techniques presume that a reasonable, well-fitting relationship exists between the prices of housing feature (or any goods) and the characteristics of those features or goods. This relationship can be expressed as:

$$V = f(C_1, C_2, \dots, C_n)$$

where V = price (or value) of the house, and C = characteristics of the house. The exact relationship between housing characteristics and price

Table 4.4 Comparison of Housing Options in Terms of Economic Rate of Return, Net Present Value and Net Present Value/Total Cost. San Salvador 1978

Housing option	Rate of Return	Net present value (Colones)	NPV/Cost	Ranking on 3 indicators (1 = highest)
<u>Upgrading and sites and services</u>				
FSDVM Basic Unit	33	4065	1.2016	1
FSDVM Serviced lot	28	2329	0.7269	2
IVU Rehabilitation	18	1078	0.2640	4
<u>Traditional housing</u>				
IVU Multifamily units	9	-1828	-0.1304	9
IVU Single family 2 bedroom unit	11	-606	-0.0720	8
FSV Single family unit	13	452	0.0641	5
<u>Informal market</u>				
Colonia ilegal	22	1788	0.3500	3
Heson	12	1674	0.0141	7
Tugurio	20	373	0.2972	6

Source: Marisa Fernandez-Palacios and Michael Bamberger "An Economic Analysis of low-cost housing options in El Salvador." DEDRB (draft) August 1979.

is not known. It is assumed, though, that this functional relationship can be expressed in the linear form:

$$V = P_0 + P_1C_1 + P_2C_2 + \dots + P_nC_n + \text{error terms}$$

and that this equation can be estimated using linear regression analysis. The coefficients of the characteristics can be interpreted as the shadow prices of those characteristics. For example, if C_1 denotes the number of rooms, P_1 measures an additional room's contribution to the total price (or value) of the house.

Hedonics have been useful in measuring changes in housing quality over time. In the Philippines, observations of housing characteristics were taken twice, before structures were affected by reblocking, and again six months after reblocking. Data from the first period, containing information on the houses' estimated values, are used to estimate a hedonic equation of the form:

$$V^0 = p_0^0 + p_1^0 C_1^0 + p_2^0 C_2^0 + \dots + p_n^0 C_n^0$$

where the superscripts refer to the initial period. The above equation is estimated for the full sample, and assumes the p_1 , for example, captures the market price of the first characteristic. V^0 gives the estimated value in the initial period of a house subsequently affected by reblocking.

In the second period, housing value would be V^1 , which, strictly speaking, would not be comparable with V^0 as a measure of quality change, since prices may have altered in the meantime. Because

housing prices are hard to measure, due to the indivisibility of the good, housing cannot simply be "divided out." However, if the estimated hedonic coefficients from the above equation are substituted into another equation using characteristics in the second period, one can get an estimate of V^1 in terms of the initial period prices, or:

$$V^1 + p_0^0 + p_1^0 c_1^1 + p_2^0 c_2^1 + \dots + p_n^0 c_n^1$$

The relative change in housing quality can then be estimated as

$$Q = V^1/V^0 .$$

The actual measures used in these equations in the case of the Philippines were drawn from two sources: 1) estimation by 96 household heads of their structures' worth and the value of individual housing characteristics (floors, walls, etc.) and 2) valuation of the structures by a trained independent appraiser, contracted to the National Housing Authority (NHA). With regard to housing value (See Table 4.5), results indicate that the owners' assessments of their houses' worth (average 14,145 pesos). In addition, the two estimates are highly correlated with one another and yield similar results when used as dependent variables in hedonic price equations. The determinants of squatter dwelling value tend to be similar to those for formal sector dwellings. A house's external appearance, and the type and quality of materials used in its construction are among the most important variables, although the latter's significance is affected by the availability of a choice of building materials which not all project cities (such as Lusaka) offer. Lot size is another important measure of value.

Table 4.5

COMPARISON OF INDEPENDENTLY APPRAISED HOUSING VALUES
WITH OWNER ESTIMATES 1/

		Philippines <u>2/</u> 1979	Kain and Quigley 1972 St. Louis Study (All Owner Occupied Homes)
(1) Average Appraised Values	$\frac{1}{N} \sum_{i=1}^N C_i$	\$1879	\$14,431
(2) Average Owner Estimated Values	$\frac{1}{N} \sum_{i=1}^N O_i$	1886	14,473
(3) Difference [(1) - (2)]		-7	-42
(4) % Difference [(3) ÷ (1)]		-0.4	-0.3
(5) Absolute Value of Differences	$\frac{1}{N} \sum_{i=1}^N C_i - O_i $	1027	3,058
(6) Pearson Correlation Coefficient		.765	.87
(8) Spearman Correlation Coefficient		.776	-
(9) Sample Size [N]		96	113

1/ $C_i = \text{CONSVAL}_i$ = value of the i^{th} home as estimated by consulting professional appraiser.

$O_i = \text{OWNRVAL}_i$ = value of the i^{th} home as estimated by the owner of the i^{th} home.

2/ Table 1 figures converted to US dollars at US\$1 = P7.5 Philippine pesos.

Standard errors are in parentheses.

Source: Emmanuel Jimenez, "The Value of Squatter Dwellings in Developing Countries," Urban and Regional Report No. 80-17, DEDRB, World Bank, 1980, Table I, p. 19.

Interestingly also, there is preliminary evidence that a structure's age correlates positively with value. ^{1/} It may be inferred that a certain age demonstrates a house's durability: visible evidence of the resources invested through progressive development, as well as a sign that longevity in a particular area suggests low risk of eventual displacement. The average structure in Tondo, for example, is 12 years old, a measure of Foreshore residents' capacity to withstand numerous threats of eviction and razing.

Surprisingly, it appears that water and sanitary facilities are not valued as much in the Tondo project as in the market. This finding may accurately reflect the priorities at given income levels and stage of upgrading, which may be exacerbated by the installation problems and delays experienced. On the other hand, it may provide an indication that residents are not yet acquainted with the health benefits of toilet sanitation. If that is the meaning of participants' relative disinterest in water and sewerage facilities, a project incorporating such services could include educational components to alert project dwellers to their hygienic advantages.

Examined over time, most measures register some sort of improvement in housing quality (see Table 4.6). The changes are especially evident in the greater proportions of households with solid walls and concrete foundations. Lot and building areas generally

^{1/} More detailed discussion of the hypotheses behind these measures and of the estimation results may be found in Emmanuel Jimenez, The Value of Squatter Dwellings in Developing Countries, Journal of Urban Economics, 1982.

increased also, as did the average number of floors. The proportion of structures with water connections declined slightly, probably due to delays in providing water. The hedonic equations reveal that a composite indication of housing quality in Tondo has increased from 60 percent to 85 percent. In monetary terms, the absolute increase in housing quality is estimated to be valued at between 6200 and 8000 pesos (approximately US \$800 to \$1000). Within a short period, the project has stimulated housing investments that in turn have raised overall dwelling quality substantially.

Still, not all households have indicated increases in housing quality since reblocking. Given the relatively short time span in this case between reblocking and the second series of observations, some 34 to 37 percent of families had been unable at that point to upgrade their houses to pre-blocking levels. Considering the overall magnitude of improvement, then, the 63 to 66 percent undertaking investments apparently did so on a large scale. If it not yet clear how socio-economic characteristics intersect with these results, currently a topic of further research. It is evident though that improvements are still being initiated, and that squatter housing consumers behave in an economically rational way, valuing dwelling units similarly to conventional markets.

Some measures of housing quality were sought in Zambia. The results were high, relative to conditions in African countries, at least by informal measures. Walls are constructed of concrete block or burned brick in 89.5 percent of the Matero houses, of concrete block in 92.5

Table 4.6

HOUSING CHARACTERISTICS IN THE TONDO AREA 1/

<u>Variable</u>	<u>Description (Averages)</u>	<u>Mean Before Reblocking for Affected Sample 2/</u>	<u>Mean After Reblocking for Affected Sample 2/</u>
AGE	Age of the structure in years	8.58 (4.25)	10.2 (4.45)
CMNTWALL	Proportion of dwellings with solid (cement or brick) walls	.16 (.37)	.47 (.51)
FINWALL	Proportion of dwellings with wall finish (e.g., paint)	.03 (.28)	.05 (.16)
SOLIDE	Proportion of dwellings with concrete foundations	.08 (.27)	.26 (.45)
LOT	Average lot size in square meters	61.3 (64.5)	72.2 (146.3)
BUILD	Average building area in square meters	32.1 (16.3)	53.4 (15.3)
STORY	Number of floors	1.4 (.50)	1.6 (.50)
TOILET	Proportion of dwellings with bucket-flushed or other water-sealed toilet	.29 (.46)	.50 (.51)
WATER	Proportion of dwellings with sink (and water connection) installed	.92 (.27)	.84 (.37)
RICH	Proportion of dwellings in neighborhoods (super-blocks) with monthly average incomes above 1,000 pesos	.26 (.45)	.26 (.43)
	Number of observations	38	38

1/ A six month period elapsed between the two measurements.

2/ Standard deviations in parentheses.

percent of the Lilanda houses, and of sun-dried brick or concrete block in 91.4 percent of the George houses. Roofs are made of asbestos cement in over 95 percent of houses in Matero and Lilanda, and of corrugated iron in over 87 percent of George houses. As for plot allocation, over 96 percent of residents in Lilanda and Matero were satisfied with their plots. Regarding infrastructure, over 71 percent of Lilanda residents thought it was of good quality. Over 92 percent believed that on an absolute scale their overall housing quality was good, with less than 1 percent thinking it bad. While these percentages decline somewhat when beneficiaries are asked to compare their housing with other types provided by the Lusaka City Council, over 75 percent still believe their housing to be better than that in other LCC projects.

Evaluation findings show that housing is considered to be an investment as well as a consumption good, and suggest that project designers must take this into account. The shelter programs originally conceived in the early 1970s were intended to increase the consumption of housing, viewed as a basic need undersupplied in most developing countries, rather than to focus on housing's potential as an investment. In order to ensure that shelter benefits accrued to the low-income target populations, certain restrictions on use and disposal of the properties were applied. For example, the Tondo upgrading project prohibits resale of developed plots for five years, in the hope of minimizing speculative encroachment on the project by higher income groups and by absentee landlords and developers. In El Salvador and Zambia, the renting of rooms was also prohibited, with the additional

objective of limiting densities and thus preventing the deterioration of health conditions. There is no doubt that these restrictions on ownership rights, affecting development, use and resale must have repressed investment demand, how much depending on how consistently the restrictions have been observed. These questions are the subject of current research.

Studies of these and other recent Bank projects have provided evidence that housing is indeed a profitable investment, constituting a major outlet for private household savings, generating employment at low foreign exchange cost, and yielding a flow of income. The previous chapter discusses the marked difference in propensities to consume housing between renters and owners. Although the attempt to quantify the "causes" of this difference is only recently underway, it is obvious that some of the difference is accounted for by the construction of additional space or rooms for rent. This inference has been substantial for the control areas in Senegal, and is clearly an important factor in other markets as well. The very high proportion of houses or parts of houses that are rented in the colonias ilegales of El Salvador, shows that the investment motive is clearly operative there.

The evaluations have found that renting is one of the most effective ways of increasing income, and hence of making or keeping projects affordable. This finding holds true particularly for participants in the lowest income strata. A related observation, now under study, is that rent and other means of income generation through housing (such as industrial or commercial use of part of the structure)

are frequently dependent upon a beneficiary's ability to coax transfers from the extended family to finance initial construction. The poorest 20 percent of Zambian project participants derive about 25 percent of their total income from rentals, whereas only 5 percent of total income is drawn from rents by the wealthiest 10 percent. It should not be inferred that renters are destined to supply an ever-increasing proportion of owners' housing costs, as there is no evidence to date that the projects are helping to create or enrich a rentier class or that rents have been rising since their construction. Preliminary evidence from the Philippines suggests that, although nominal rents have increased by 11 percent since project initiation, given inflation, this represents a decrease in real terms. The incidence of benefits does not seem to be shifting from participant families to non-project families, though there is still concern that affordability problems may stimulate such a course of events someday.

Access to Services

Attempts to make projects enhance access to social and other urban services have so far shown mixed results, although in all instances environmental quality and integration with urban services have clearly improved.

The most thorough effort so far to assess project effects in this sphere has been in El Salvador. There the relative scarcity and high price of land compelled the FSDVM to situate projects in peripheral locations. The data indicate that project households in Santa Ana are within 600 meters of school, public telephone, and park or playground,

and that virtually all are within 200 meters of public street lighting. Medical assistance and public transport are on average only 1000 meters distant, and for most people the local market and their place of work are within 2000 meters. This means that for most people all the basic services and place of employment are within a half hour's walk of their home. These distances are relatively small, although it must be remembered that, like most Salvadorean cities other than the capital, Santa Ana is also fairly small. Thus residents in mesones are on average 60 percent closer to basic services than are project participants, whereas families in tugurios are on balance equidistant. Those in colonias ilegales are on average 20 percent further away than project dwellers.

Table 4.7 shows the effect moving to the project would have on access to services for families presently living in mesones and colonias ilegales. The meson family would on average have to travel an extra 300 meters to reach these services. The greatest increases would be for visits to the nearest market (+1450 meters), a health center (+400 meters), public transport (+850 meters), and place of work (+500 meters). The only substantial improvement in access resulting from a representative move from meson to project would place families 400 meters closer to a park or childrens' playground.

From the viewpoint of families presently living in colonias ilegales, a move to the project is more attractive in terms of proximity to services. There the greatest improvements of access involve childrens' playground (-1100 meters), public telephone (-650 meters),

DISTANCE TO PUBLIC SERVICES AND WORK IN PROJECT COMPARE
TO MESONES AND COLONIAS

Table 4.7

4-31

Difference of Mean Distances
Mesones Colonias

Medical Aid	+ 407	- 460
School	- 130	- 240
Transport	+ 844	+ 673
Public Lighting	- 28	- 210
Public Telephone	- 95	- 638
Market	+1453	+ 518
Playground	- 420	-1090
Work	+ 504	+ 419
Average	+ 309	- 137

and medical assistance (-450 meters). On the other hand, the project here too is further away from the public transport (+650 meters), market (+500 meters), and place of work (+400 meters).

In general, these differences are not excessively important, as in most cases the distances involved are very short, involve trade-offs and such findings may also be soon obsolete. It is worth noting that there was a trade-off in El Salvador between improved sanitation and access to a childrens' playground, on the one hand, and distance from work, particularly for people who are self-employed small traders, on the other. One advantage of the meson is that a family can operate an informal sector business from the house, so that, for example, a mother can manage a small store while looking after her children. A move to the project, even though it does not involve a great increase in distance, may mean a loss of "access" to the customers who frequent the stores in the busy town center where the mesones are located. On the other hand, it is to be expected that the main distance disadvantages between project sites and either mesones or colonias -- notably distance: to market and to public transport -- will not persist for long. The large number of small shops which quickly spring up in project sites attest to that.

Similarly, it is predictable that the dynamism of the communities will serve to reduce differences in acces through

transport, ^{1/} such that remaining differences in access to services is determined by other factors, as well as distance. In the case of water supply, many families have water available to them only at certain hours of the day, or have to spend time queuing. Neither of these problems exists in the project, this being an area in which a move to the project results in an important and unambiguous increase in access. In colonias, only 62.2 percent of families with individual water connections have a water supply 24 hours a day. For mesones, the figure is 84.5 percent. For those families that have to purchase water in tugurios or colonias, virtually all have to spend at least one hour per day queuing and carrying, and almost half have to spend two hours or more.

In general, project participants seem to be quite satisfied with access to schools, water, and public lighting, though most are dissatisfied or only moderately satisfied with access to medical services and public transport. Table 4.8 presents a comparison of levels of satisfaction on these variables among project participants of families in other informal housing. Project participants rate access to medical services slightly below average, and access to public schools and water about average. The only area in which project participants are more than averagely satisfied is public lighting. (This must be recording mainly a quality or reliability difference, since Table 4.7 shows that "access" has improved only marginally in terms of distance.)

^{1/} If this does not occur, it will likely be due to the inefficiencies in the public transport sector, or restrictions placed on the private transport sector.

Table 4.8

MEAN SATISFACTION WITH ACCESS TO SERVICES. COMPARISON OF PROJECT, COLONIAS,
MESONES AND TUGURIOS. SANTA ANA 1980.

(In Percent)

	<u>Participants</u>	<u>Mesones</u>	<u>Colonias</u>	<u>Tugurios</u>	<u>Mean</u>
Medical Services	40.1	77.6	64.8	68.0	59.0
Schools	97.0	95.0	86.0	88.0	92.4
Water	93.0	95.0	87.0	87.0	90.9
Public Lighting	97.0	92.0	35.0	40.6	71.2
Public Transport	14.0	97.0	78.0	86.9	60.0
Mean	68.2	91.3	70.1	74.1	74.7

Trade-offs between project benefits are an inevitability. The project provides higher standards of water supply and sanitation, but is further from employment and other facilities in the central city areas where the mesones are located. For a self-employed small trader, proximity to place of business is important, but for most other groups the small size of the city have means that the higher level of services provided in the project can more than offset the relatively small increase in distance to some public services.

Experience to date in the three other project countries adds two important findings. First, provision of all-weather road access is a very important factor. In Zambia, installation of roads has led to the project area's businesses being better supplied, and demand for transportation has exceeded public facilities' capacity, prompting private sector employers to provide transportation for their employees. In the Philippines, roadways within the project areas have been greatly upgraded. Even at this early date, traffic has increased considerably, and prospects for further development seem good.

The second finding is that delays in the provision of participants' access to certain facilities can have important adverse effects on subsequent steps. In Tondo, delays by the agency charged with supplying water resulted in participants making only slight improvements to sanitary facilities in the period immediately following reblocking. Delayed access to services has been one of the major problems of the Dakar site, which is located seven miles from the downtown area. Problems still exist with the provision of electricity,

public toilets, schools, markets, and security arrangements, and continue to retard occupation by householders and the development of businesses. In Zambia, dissatisfaction with street lighting and garbage collection have contributed to cost recovery problems, as has the policy of cutting off water supply of groups of families who share standpipes in an effort to enforce collection from individuals.

Employment and Income Generation

Projects have sought to stimulate employment and income generation in three ways: through the use of hired labor during house construction, ^{1/} through specific, project-designed employment and business components, and by means of induced effects. The operative assumptions are that the opportunity to own a better, more secure home will stimulate families to offer more labor, and that improvements affecting the project will bring families' labor and employment demands closer together. The evaluations have found to date that it is the first method that has produced the most clear-cut results. In the El Salvador project, housing construction yielded \$4.16 million in wage income, and 3700 person-years of employment, compared to original estimates of \$3.5 million in income and 3500 person-years of work. Some 8000 person-months of labor and over \$1 million in wages have been generated in the Zambia project, which contains significantly smaller sites and services components than the Salvadorean project. As

^{1/} This, of course, is an alternative to use of family labor for self-help and mutual-help for families that find it rational to hire labor rather than expend their own alone. Seeking remunerative uses of their own skills thus enables them to pay for the skilled labor necessary for efficient house construction.

mentioned, Zambian families, chiefly female-headed households, have increased their incomes by as much as 25 percent by renting space in their new or upgraded accommodations.

In the ongoing projects, statistics to date reveal merely that labor is being hired to help complete housing construction. Preliminary data from the Philippines show that the ratio of the number of paid to unpaid person-days of labor is approximately 2:1, with some 30 percent of households relying exclusively on hired labor and another 49 percent supplementing their own unpaid labor with some hired labor. In Senegal too, it has become clear since the end of 1977 that households prefer to hire small contractors to aid in housebuilding, a principle adopted and developed by project management since 1979. At the end of 1980, seven supervised contractors were at work on the site, and over a period of six months had completed 118 two-room core houses, at the moderate sum of CFAF 85,000 (\$) per dwelling.

The expectation of some induced employment and income generating effects was vaguely written into the socio-economic justification sections of some appraisal reports. This was done in spite of the early assumption of near-zero opportunity cost of participants' own labor. Though they were not clearly formulated at the outset, the conceptual justifications for this expectation can be understood as follows: the opportunity to own an asset--for many potentially an earning asset--that can be progressively improved, and whose value grows through improvements and through increases in the value of its location, stimulates investment. This investment will be

supported, perhaps with a lag when borrowing has been feasible, by increased income earning activities of the household. For some households this will mean increased average labor force participation, which will be facilitated by the improved quality and services access of the neighborhood.

The evaluation effort has provided some evidence of the correctness of the project designers' expectations: the high proportion of hired labor used in construction (59 percent in El Salvador and roughly two thirds in the Philippines) suggests the willingness and capacity of at least some households to expand their labor market earnings. Studies of incomes, expenditures, and affordability have shown that housing expenditure/income coefficients are greater for "unearned" income than for wage earnings, and greater for earnings of secondary workers than for those of the household head. Finally, there is some suggestion from studies in El Salvador that labor force participation and earnings improve for participant families, relative to controls, specifically among secondary workers. Further estimation with additional years' data will be needed to determine whether this finding describes a stable relationship or a temporary phenomenon. We won't make much headway in this area without a carefully articulated model of the household economy, such as Jimenez has developed. ^{1/}

Regarding specific employment components, as many as 14 cooperative organizations have been formed in El Salvador to create

^{1/} See Emmanuel Jimenez, "The Economics of Self-Help Housing: Theory and Evidence from a Developing Country," Urban and Regional Report 80-16, DEDRB, World Bank, 1980.

employment, including a building materials shop, a bakery, and a dress and rug-making craftshop. The first two have proved economically viable, though with little growth potential, while the third encountered serious administrative problems. Because only a small proportion of all project participants are able to work in these enterprises, they have been viewed as an attempt to foster an economic elite, and have hence been unpopular. Other ongoing cooperatives, aimed at savings and loan arrangements or at satisfying widespread consumer needs, may prove to have a broader appeal and be more acceptable in future projects.

Out of 100 commercial plots advertised in three areas in the Zambia project, only 31 applications had been approved as of March, 1979, of which 10 had led to completed structures, and another 13 were under construction. The reasons for this lack of response seem to have been the absence of business loans to support the component; high building standards discouraging many people, especially those without access to materials stores; remote plot locations making access difficult for clientele and suppliers; and advertising campaigns that were insufficiently detailed to stimulate interest.

In the Philippines project, five different employment generation efforts have been introduced. A first program provided some 500 residents with jobs in the construction firms. However, because this program was not monitored, little is known about the kind of work performed or its duration. A second, small business loans component was originally not well publicized, and the associated application procedures seem to have been unduly cumbersome. Over 230 loans have

nevertheless been granted, largely for manufacturing and trading activities, and a slowly expanding stream of applicants and workers is being brought into the program. Cost recovery problems in the loans programs have been significant, however. A third, vocational skills program, while well-executed, has not led to large scale job placement. A fourth, cottage industries program, including shellcraft and conversion of garment scraps into toy and pillow stuffings, has proved to be unmarketable. Nonetheless, popular interest in new ventures continues to be high.

The larger-scale commercial/industrial estates in the Philippines have also failed to provide jobs in the numbers projected or to make rental payments on their buildings when due. The National Housing Authority, which has overseen the component thus far may yet seek technical assistance to help the estates operate more efficiently. ^{1/}

Finally, expenditures for food, medicine, and other basic necessities do not seem to have been adversely affected by participants' attention being focused on housing, as some had feared. In Santa Ana, El Salvador, for instance, project participants registered a greater increase in per capita expenditure on housing than the control group, but there were no observable differences between the two groups' changes in expenditures for food, medicine, and transport. In Zambia, plot size has permitted families to satisfy some basic needs directly by growing

^{1/} "Interim Report," Tondo Foreshore Dagat-Dagatan Development Project Report Series 80-1, Research and Analysis Division, National Housing Authority, 1980, pp. 38-44.

their own food, and there may well be a basis for stimulating such activities there and elsewhere. Residents surveyed believe they could not satisfy their food needs without being able to grow crops themselves. Other studies of the issue have cited that urban agriculture can free households from the vagaries of international food markets, allowing city dwellers to supply their own consumption needs and seek small profits through the sale of surpluses. The practice can also give African urban women a measure of autonomy not achievable in other domains.

Impacts on Urban Housing Policies and Urban Areas

The most important result of project activity in this regard has been the demonstration that low-cost housing can be built and made affordable to the urban poor. As much as 70 percent of the population of a normal upgrading project in Zambia is comprised of members of the poorest 30 percent of the urban households. Even the more complete houses constructed in the Zambian overspill areas cost under 20 percent of the cost of the cheapest contractor-built government housing. Upgrading as a form of progressive development has been fully incorporated into that country's Third National Development Plan, and a new effort in Kalingalinga, funded by the German government, follows the model that guided the Bank's Lusaka project.

In El Salvador, 14 to 18 percent of the third, fourth, and fifth income deciles of the total urban population of the nation's secondary cities have been successfully incorporated into the project. Again, even the relatively more elaborate housing built in the program

cost less than half as much as that constructed under other public housing schemes.

A second impact has been the general recognition that, whilst low-cost shelter can be provided, such projects make heavy demands on cities' capacities to provide services in the upgraded or newly developed areas. These demands range from the lengthy process of legitimizing purchases of land to the actual provision and maintenance of water, lighting, and sewerage and garbage disposal services. Although difficulties in meeting these demands imply temporary limits to the scale and complexity of projects that might be implemented, more dynamic policies can be formulated over time and in the light of citywide pressures to deliver basic services. Problems with cost recovery in the Bank's (and in other) Zambian shelter projects, for example, point to the need for examination of a full range of factors - including massive housing subsidies at all income levels - underlying the phenomenon. Use of a public finance framework, encompassing more equitable systems of charges for housing and urban services across income strata is proposed as an aid to the projects' paying for themselves, and to generate additional revenues for maintenance and further improvements. The Bank has already begun to design interventions that integrate housing programs with improvements in urban utilities and fiscal policies.

Another, related category of impact is the projects' influence on the nature and direction of urban growth itself. Preliminary studies in Zambia indicate that a greater integration of richer and poorer

neighborhoods has come about, with residents cooperating for achievement of mutually held objectives. The general sentiment among project officers is that the project has increased political awareness in Zambia without inflaming political tensions, a notable accomplishment, given the highly-charged, dynamic atmosphere of the squatter communities at the outset of the program.

Comparable social results have been observed in the Philippines project. The Tondo area has clearly benefitted from legitimization of its land claims. The disruptiveness implied in the reblocking effort has not led to increased social instability. The barrangay appears to have been strengthened as a socio-political unit responsive to development efforts. Cooperation among families in the project has been evident, and crime in the area appears to have dropped, according to informal surveys of police activities. On the other hand, initial impressions have been that social stratification in the Tondo area has visibly intensified, with differentiation made evident by conspicuous consumption and investment in housing properties. Though this consequence may not be economically negative, expansion of the wealth differentials implied may foreshadow future exclusion of less affluent families with unwelcomed social repercussions. This tendency will require continued study, already proposed for the Philippines context.

Chapter 5:

EVALUATING PROJECT EFFICIENCY

Although evaluation research to date has affirmed that the progressive development model is viable, and that project effects and impacts have occurred more or less as planned, there are indications that projects have not performed as efficiently as expected. To a degree this outcome appears due to the heterogeneity of the populations found in squatter communities. Heretofore thought of and described as low-income areas, they have been found to contain large concentrations of the relatively poor (e.g., 30 to 40 percent in the lowest pentile in Zambia and the Philippines), but also a fair representation of all other income strata except the highest (e.g., 50 and 40 percent, respectively, in the upper six deciles). In terms of income characteristics, the population are best described as middle income. In terms of housing conditions and demand, they are probably best characterized as low rent areas, or areas where households (whatever their incomes) seeking low rents tend to settle. ^{1/} These facts limit to some extent the possibility of improving targetting on the poorest through upgrading projects. But it is quite clear that certain adjustments and changes in project design could improve delivery to target populations in sites and services and in some upgrading situations.

^{1/} It is unclear whether the process by which "middle income" households wind up in such areas is active (i.e., they move in) or passive (i.e., they move in while poor but tend to stay even after an increase in income). Probably both types of behavior are common, and it is not yet clear that it matters which predominates.

This chapter shifts focus to the question of how efficiently Bank sponsored projects meet their objectives of delivering low-cost, affordable, and user-acceptable shelter solutions expeditiously and on budget to intended beneficiaries.

Although differences exist in the goals and operational procedures of the four shelter programs reviewed in this book, there is enough basis for comparison to warrant investigation of their efficiency according to a single set of criteria. Eight project elements appear to be of particular financial, economic, and social import in all four programs: 1) project planning and design; 2) selection of project beneficiaries; 3) construction methods; 4) materials loan programs; 5) housing completion and occupancy; 6) maintenance of housing and infrastructure; 7) cost recovery; and 8) community participation. The following criteria are most useful in gauging the efficiency of these components: speed of implementation; cost; quality (of housing or servicing); accessibility to target population; replicability in different national contexts; and flexibility of implementation.

Inevitably there are trade-offs among evaluating criteria. Higher quality of house construction may entail higher costs and less accessibility to the poorest. Greater attention to accessibility will usually slow the process and increase its cost. There is no one ideal solution, for the problem is to balance many mutually impinging factors. It is necessary to try to reconcile sometimes conflicting objectives according to the relative weight of each as determined by project management and higher level policy makers. The following

discussion summarizes the constraints and options involved in applying the criteria to each project element and reports initial results from evaluation research.

Project Planning & Design

Two principles are critical in this area. First, design standards should be of a level and degree of flexibility that neither discourages participation by the target group nor impedes implementation. Second, the speed of implementation can affect the cost of both services provision and house construction. If house consolidation proceeds too slowly, administrative costs will rise and intended benefits will be delayed. Cost escalation will be exacerbated during inflationary periods. Pressure to speed up house construction, however, may lead to reduction in a family's use of its own labor and thence eventually to affordability problems.

The principle causes of implementation delays have been difficulties in acquiring land and problems, particularly in relatively complex projects where substantial coordination is required, with the performance of important associated agencies. If delays of either kind are serious or sustained, they will create difficulties in getting households to move to the sites, thus impeding the house consolidation process still further. This was the experience in the Senegal Project. Since such delays have cost consequences, they raise distributional as well as efficiency concerns.

It is not possible at this point to make confident assertions about appropriate design standards. However, findings with respect to

affordability and accessibility indicate that they could be lowered -- or at least more lower cost options provided; This could improve targetting without apparently compromising other objectives.

It is true, as well, that families in the lowest pentile of income distribution are under-represented in the projects, particularly in the El Salvador Sites and Services projects. However, it has not yet been determined that this was the result of low-income families deciding not to join the projects because standards and costs are too high. There is evidence that low income has not so far emerged as the factor to explain the withdrawal of families from the projects.

We also have evidence from all countries that project designers have occasionally erred in their judgments of what project features will be favored by participants, particularly low-income ones. There has been a general tendency to overemphasize certain services and community facilities (notably electricity, water supply, and sewerage) and to underemphasize housing space. This was first apparent in Africa, where, in Senegal, it was found that the control populations tended to build larger houses than expected for the project and to sub-let part of the space. At the same time, they were slow to install electricity and plumbing, in part because connections were costly and difficult to arrange. Water-borne sewerage was not offered in the Zambian project. However, the desire to trade off other project features against more space was recorded there too.

These tendencies were at least as pronounced in El Salvador and the Philippines. In the former country, demand studies in the smaller

interior cities recorded a substantial proportion of low-income families preferring the larger plots and lower service levels of illegal subdivisions to the FSDVM options. Cost-benefit analyses confirmed that the two options were approximately equally attractive. In the Philippines, the preponderance of multi-storey houses again points to the substantial desire for space; hedonic analyses placed a second storey near the top, and sanitary facilities near the bottom, of the spectrum of components evaluated.

What these findings seem to indicate is that when projects offer only limited options and flexibility, the cost of guessing wrong is substantial. Many of the intended beneficiaries will find that project options do not suit their requirements. This has been particularly true in regard to provisions for rental accommodations. These have either been missing, or, in the early projects, proscribed, stifling what has proved to be a considerable demand on the part of the target populations. At the same time the desire of others within the target population to supply rental accommodations to their fellows has been stifled by project limitations, especially in El Salvador and Zambia.

Another question raised by project experience is the extent to which community facilities (such as school, health centers, parks and playgrounds, etc.) should be designed into a project or provided for as future possibilities. The desirability of such features is acknowledged, and has been demonstrated. On the other hand, each new facility brings at least one more governmental agency into the picture,

with all the additional demands for coordination, scarce personnel, etc. that this implies. Complexity contributes to delay, ^{1/} and it is worth considering whether such components should not be deferred until they are demanded by participating households through community political organizations (with more done to encourage this organization). They might then be supplied more efficiently by larger sectoral programs.

Selection Procedures

Studies to date have not revealed any important anomalies or irregularities in project selection processes or their results. In fact, in El Salvador and Zambia, where the genuine concern of project administrations to reach the lower-income groups predates involvement by international agencies such as the World Bank, selection procedures have had a very good record of admitting families within the prescribed income ranges. The apparently contradictory indications of Table 4.1 remain to be explained.

First, with respect to the squatter area upgrading programs, the problem is not one that can be resolved by altering the selection processes. The character of these settlements is different from what had been envisioned, and a project has to deal with a population already present, rather than with one that will effectively select itself by its desire to move in.

While it is necessary to work with a population already there in squatter areas, the selection processes remain useful means of

^{1/} Which is germane not only to immediate projects, but also -- and particularly -- to the program's chances for success.

determining beneficiaries of sites and services schemes. To date, evidence shows that these have worked reasonably efficiently and equitably, if not perhaps as aggressively as they might have. There seems to have been some tendency for applicants to distort their incomes (downward in the upper ranges) to ensure their entry into projects. At the same time, there was a tendency to set project income ceilings higher than was desirable from a distributional perspective, in order to ensure that early projects would not be undersubscribed. It is not clear how prevalent these practices were nor how serious have been their consequences.

Adjustments of affordability criteria, such as the consideration of total household income, rather than wages of the household head, will reduce remaining biases against sub-groups of the target populations. In El Salvador, for instance, just the addition of transfers from the extended family into total income increased the incomes of many poor families by as much as 20 percent. A general finding has been that some families, particularly female-headed households and those employed in the informal sector, may have suffered some unintended bias in projects' initial stages, due principally to problems in verifying income from non-formal sources. This was first observed in the early stages of the Dakar project, and has been corrected in some measure. In El Salvador there has been a concerted effort to avoid such discrimination, and projects have gone so far as to include a higher proportion of female-headed households than is found in the population at large.

The generally successful FSDVM selection procedures may serve as a model for other programs of their kind. A good indicator of the effectiveness of these procedures is the fact that, once families complete construction of their houses, the drop-out rate is extremely low, the strict enforcement of collections demonstrating that all families have the capacity to pay and are genuinely interested in staying in the project. Still, even in this very successful program, there are some unanswered questions. For example, it is not clear what can be done about one major category of drop-outs -- families who withdraw during the selection process and who cannot be relocated. These families are simply lost, even though eligible.

There has been a continuing debate over whether participation in mutual help construction should be necessary requirement for selection. Whilst the whole logic of the FSDVM program is based on community participation, there is strong evidence that eligible families may be excluded by this requirement.

The cost of the FSDVM selection process itself is an issue. Considerable resources are devoted to it, particularly in the initial orientation and personal contact stages. Selection is thus a long and arduous process, with high drop-out rates all along the way.

Finally, despite the increasing scale of projects, there has been resistance within the FSDVM to computerization of the selection process on the part of those who feel that the human element of the development process would suffer. This concern arises from the fact that the personal judgment of the promoter plays an important part in

selection. There are ways that the judgment of the promoter could be incorporated into a computerized selection process without unnecessary compromise. For example, the promoter's opinion of a family's eligibility could be coded, and selection made in terms of the normal weight of criteria. Individual attention could then be given to all cases in which the promoter's opinion was inconsistent with the recommendation based on the weighting system. However, persuading applicants that the human element is being attended to might be more difficult. For this and other reasons the issue of computerization becomes more pressing as the size of projects increases.

Construction Methods

Project designers have made a priority of identifying those construction methods that would reduce costs and contribute to efficient implementation. Among these are construction through mutual help; self-help in which families hire a contractor; self help in which families themselves contract, hire, and supervise workers; self help in which families build their homes themselves; and project-provided, and contractor-built, housing. Each method involves different rates of construction, different costs, and different implications for housing quality, replicability, and accessibility. The mutual help method tends to be slower (though not necessarily less expensive) than certain self-help methods, which may also discourage participation by female-headed households, as women would be expected to help in the construction process just as men do, but would be less inclined to do so. The timing and flexibility provided for self help and mutual help methods can also

encourage or discourage families with varying formal or informal work schedules. Similarly, housing quality will be affected by the amounts of supervision or technical assistance mandated by project authorities. Use of skilled builders can raise project costs in the short run, but their help in project design might also contribute to replicable plans that can scale down costs over time.

Early studies of the mutual help process revealed both advantages and drawbacks. This procedure was used for housebuilding only in El Salvador. Most disadvantages that could be eliminated without altering the essence of the approach have been, and the process appears reasonably efficient in producing housing of similar quality to that produced by contractors, at comparable or lower cost. The apparently irreducible problems of the approach are that administrative requirements are considerable, and that the implementation time must be extended, because groups can be gathered to work only at certain times. Presently, most such work takes place on weekends, though it is not clear that this is the only time, or even the best time for such work to take place. The matter is still under study. In some circumstances, these factors may make the approach unviable. In El Salvador, they do not appear to do so. There the advantages outweigh the disadvantages, including enabling participants to create "sweat equity"--a downpayment on their houses--through labor.

"Sweat equity," however, does not appear to have been available -- in El Salvador or the other countries -- in the abundant supply assumed, whereas financial resources, especially income

transfers from relatives, are more available than anticipated. Surprising early findings concerned participants' opportunity costs: the first was that for many, notably those in informal sector work, the weekend is the worst time to organize mutual help activities. Sixty percent of participants working on the El Salvador project reported that weekend work conflicted with other money-making activities; forty five percent of these reported that participation reduced their incomes. ^{1/} The second is that the burdens are shared quite unequally. The third is that, for the group as a whole, average opportunity costs and the wage paid by the Foundation were approximately equal, which stands in sharp contrast to the original assumption of zero opportunity cost. Although amenable to making some further improvements in the organization and scheduling of work, and with respect to participation by various family members, the Foundation has no intention of making other changes. The reason is that the Foundation values highly the shared work experience as a factor leading to superior cost recovery performance, future community development activities, and participation in self-governance.

Family based construction methods using progressive development have proved to be viable in sites and services and upgrading projects. It has been estimated that certain families in El Salvador have saved up to 30 percent of normal contractors' costs by building their own houses, houses that have proved to be of comparable quality to those produced by skilled builders. Still, there has been a greater

^{1/} Fundacion Salvadorena de Desarrolo y Vivienda Minima, "Evaluation of Mutual Assistance and its Functions Within the Process of Social Change," Report No. 12, July 1977, p.36.

reliance on hired labor than expected. Up to 51 percent of households in the Santa Ana and Sonsonate projects used hired labor alone to bring their core housing up to habitable levels. (Table 5.1 presents a first comparison of housing construction methods used in El Salvador.) This circumstance is not to be regretted, as it almost certainly permits faster construction and assures reasonable quality. Further, evidence is accumulating from studies of self-help, of opportunity costs (here) and of the effects of project participation on labor force participation, employment, and growth. These findings imply that "market solutions" for housing construction, such as seeking income transfers, hiring labor and "balancing accounts" through the household's own greater labor force participation ought to be considered in estimating housing costs and other aspects of project design. Self-help, it seems, should be planned as only one of a set of diverse methods for efficient housing construction and consolidation. As Jimenez' model demonstrates, ^{1/} the amount of self-help construction undertaken will depend, for each household, upon the productivity and opportunity cost of the associated activities--and hence, inter alia, upon the level and sources of household income and its skills in construction.

Materials and Credit

There is no reason to link these two items in principle. They are linked in this discussion because of their close connection in the

^{1/} Emmanuel Jimenez, "Housing Construction Methods in El Salvador," Paper for the Urban and Regional Economics Division, World Bank, 1980.

Table 5.1: HOUSING CONSTRUCTION METHODS IN EL SALVADOR:
SANTA ANA AND SONSONATE

	(1)	(2)	(3)	
	<u>Hired Labor Only</u>	<u>Unpaid Family Labor Only</u> ^{1/}	<u>Hired and Unpaid Family Labor</u>	<u>Total</u>
Proportion of Participants	.51	.27	.21	1.00
Average Number of Person- days of Hired Labor	46.0	-	33.2	42.4
Average Weekly Wage (in Colones)	124.	120.	155.	130.
Average Number of Household Members with Construction Experience	.03	.41	.12	.16
Average Number of Person- days of Unpaid Labor	-	44.3	45.0	44.6
Weekly Non-Wage Income	14.	9.	6.	10.8

^{1/} Regression analysis indicates that wage rate and proxy variables (such as technical experience, number of adults between ages 17 and 60, and the male/female mix within a household) best measure a household's ability to build its own dwelling. There is also some evidence that households who use self-help labor have lower elasticities of housing demand than others.

Source: FSDVM data.

early World Bank-financed projects. The first such project, in Senegal, did not have a credit component, in part because the project did not have explicitly anti-poverty objectives. This lack contributed significantly to project problems in execution, providing a negative illustration of the importance of credit components. In the projects under consideration, loans for housing construction have been limited to materials credits. Efficiency evaluations must gauge whether these materials have been appropriate and affordable for housing consolidation purposes, distributed speedily enough to project participats, and if other forms of housing loans might also have been advisable. There is a danger that newly created project stores will have high administrative costs, encourage overly high building standards, and fail to recommend the use of local materials, which, while acceptable for use, do not fall under the control of the stores. While replicability may be facilitated by the relatively simple "tied" credit option for purchasing materials at project stores, the technique can be compromised if the materials are too costly in themselves, or are difficult to acquire and distribute.

Although it appears that materials loans have been positively instrumental in permitting households to undertake housing consolidation, there have been difficulties in the administration of materials stores, though many have subsequently been corrected. In Zambia, for example, problems in stocking on-site stores led to considerable delays in distributing materials, causing families to have to queue for as much as 20 hours per week to receive goods. Thefts of

materials have also been a common occurrence. Restricting purchase of materials to project stores may have compelled families to buy higher quality supplies than necessary or desired. It appears that in the El Pepeto and San Jose del Pino projects in El Salvador, housing materials could have been bought at wholesale elsewhere, thus reducing the expected cost advantages of the stores' bulk purchases. The project stores in the Philippines failed to stock the cement, hollow blocks, and steel bars most desired by participants. Tondo dwellers on the whole seem to have found cheaper (yet acceptable quality) materials elsewhere, with the Housing Materials Loan Program (HMLP) providing only 25 percent of beneficiaries with construction materials. In Zambia, where materials shortages exist, families have not been able to receive the types of supplies requested. And worse, the amount of materials loans there proved insufficient to meet housing costs. The 100-Kwacha loans originally offered to Zambian upgraders proved insufficient to meet housing costs, particularly in a period of rising costs; however, instead of being increased, when the original allocation of funds had been exhausted, the loans were discontinued. Conversely, individual loan sizes in the Dakar project have been increased substantially in an effort to satisfy demand. Philippine households have been obliged to seek funds beyond the 3500 peso maximum loan available from the HMLP.

Although it is understandable that project designers were anxious to avoid having families overextend themselves financially by restricting loan funds, it seems that the "tied" credit measure has proved unduly restrictive to participants' reasonable efforts to meet

their housing requirements. This attempt to ensure that any funds borrowed go into housing (in the form of materials) rather than items for immediate consumption appears to have been misguided. In practice, the experience with respect to both materials and credit has been mixed.

It is doubtful whether materials stores can do much to expand options in countries with well-developed markets, such as the Philippines. However, they have probably made a net positive contribution in Zambia, where materials supply is restricted and many are unavailable on the private market. Even here, though, the tying of loans to materials has not prevented abuses, as some participants have sold their borrowed materials and absconded. These observations have led to a change in concept to construction loans. It has been found in Kenya, for example, that such loans can preclude most abuses by tying the release of funds to the completion of specific stages of construction.

Though analysis of the effects of tying loans to materials is not yet complete, the theory is clear as to what should be expected. Tying loans exclusively to materials will in most cases induce participants to use more materials and less labor in construction; when tied to a limited range of materials in a store it will also distort the participants' materials budgets. What may also have been important, though, is the excessive stress on self-help methods from the beginning which skewed the program and individual projects undesirably against the credit component. With the added anxiety about abuses, the project designs were steered still further away from any concept of lending for

hired labor, which currently accounts for half to two thirds of all labor used in these projects.

Most important, and perhaps most ironic, is that the credit components in most of the projects are so modest that their potential distorting effects are quite limited. Their small size seems to be the major source of constraints.

The truth is that very little is known about credit requirements of potential home-builders in urban projects, at the relevant income levels. We know very little about their demands or how these are distributed with respect to various objectives: renting, owning, generating property, etc. Nor do we know much about their access to other sources of funds, and thus about their residual demands for (potentially) formal credit, except that they vary tremendously among households. ^{1/}

Fortunately, we are accumulating valuable information on these demand questions relatively quickly. But unfortunately, we are even more ignorant about how to reduce the required amounts of credit from formal channels, how to overcome administrative barriers, and how to eliminate unnecessary and unrealistic demands for collateral, etc. Studies to date have maintained that families should be permitted

^{1/} Initial income/expenditure Studies, which have given a great deal of attention to transfers, indicate that at most income levels there are families whose access to transfers preclude the need for formal credits; similarly, there are those with very limited (or no) access to transfers. See Dani Kaufmann, "Households Income Formation and Expenditures Behavior: A Summary of Issues, Findings and Research Prospects, Urban and Regional Economics Division, World Bank, 1981.

flexibility in seeking housing materials in the market, except in those countries where materials are in acutely short supply. It is clear, though, that research into informal credit markets and their linkages with household expenditures for basic needs and into utilizing formal markets more effectively to meet residual needs, are of high priority for improving urban shelter project design.

Occupancy of Plots

The process of plot occupancy in sites and services projects is another important issue in determining the efficiency of project implementation. Unless sites are occupied relatively quickly, costs can escalate through extended interest payments during construction, through the real effects of inflation, and through the expenses families incur maintaining one residence while another is being built, or payments being made on the plot. Such cost increases may cause affordability problems for poorer families and thereby induce project managers to include more high-income families than originally desired, in order to offset these effects. ^{1/} The slower the habitation of project areas, the weaker become proponents' arguments for replicability.

Thus far, land acquisition problems have been the major source of implementation delays, and occupancy problems per se do not arise until there are plots that might be occupied. At this stage delays in providing certain basic services have been the most serious hindrances to speedy plot occupation water in one El Salvador project; water, lighting, transport, and schools in the Dakar project. Whether due to

^{1/} This happened in the Dakar project.

difficulties in alienating public land, as in Zambia, or finding non-costly and accessible terrain for lots, as in El Salvador, project managers have had to struggle literally for years to obtain land. These delays, and others outside project control, have made coordination with other agencies (responsible for water, lighting, and other basic services) more difficult. The National Housing Authority in the Philippines, for example, is obliged to negotiate with no less than 11 other agencies in order to deliver inputs in a mutually complementary fashion. Problems with credit provision have played a part too in delaying house consolidation and plot occupancy. In Senegal, the latter two problems, exacerbated by inflated expectations of housing assistance, have delayed house consolidation and plot occupancy for years. Still, the fact that families have individually and collectively found ways to surmount such massive obstacles underlines the magnitude of existing demand, as well as the resiliency of sites and services options in meeting that demand.

Maintenance

Another key concern in gauging project efficiency is maintenance of infrastructure and housing facilities. These project features frequently depend upon under-funded and over-burdened local government agencies, in a context where general poor maintenance or lack of it may be a city-wide condition. Since maintenance typically hinges on project cost recovery, and this in turn is affected by community attitudes towards the services in question, these three factors ought to be considered together in a local public finance context. More thought

and provisions will have to be devoted to these areas than was given them in the early projects, as there is considerable leeway for project planners to devise long-term financial and social solutions to maintenance problems, in conjunction with city managers and community groups.

The nature of maintenance costs is of course an important related issue. If these are too high whether because of low capital investment or poor design, the benefits of low-cost housing can be vitiated in the long run. If insufficient revenue is generated, project facilities will decline in quality. It is not yet clear whether there are feasible means whereby assignment of responsibility for maintenance to certain groups (such as higher risk households with less steady income) can be used to improve their access to projects. But it is apparent that inadequate planning for maintenance can affect government decisions regarding replicability: projects that have not helped pay for themselves---that have not produced sufficient sustained social satisfaction---are not likely to be reproduced on a large scale.

Given that these World Bank projects have been installed for only a few years, and in some cases have not yet been completed, definitive statements on project maintenance can not be made at the present time. Yet, problems with garbage collection have already arisen in the Lusaka project, due partly to difficulties in servicing the garbage trucks, and partly to poor system design that doesn't allow the trucks to penetrate far enough into residential areas requiring the service. Resulting dissatisfaction with the trash collection service

has had a negative effect on cost recovery. It is noteworthy that the local community groups in Zambia are considering handling the maintenance of communal standpipes. The FSDVM in El Salvador is also pursuing options whereby participants will accept responsibility for upkeep of certain community facilities. Because nearly all project services will require maintenance, planners will have to take additional steps in the future to ensure that both cities and their inhabitants are financially and institutionally prepared to continue handling the facilities over time, especially in the light of cost recovery problems already encountered.

Cost Recovery

Because the inability to collect payments reduces the likelihood of project's replicability, cost recovery has been deemed a crucial feature of project design. This issue, too, is complex. Delays or dissatisfaction with project services can make cost recovery difficult. On the other hand, failure to recoup payments will raise the cost of projects for the agencies involved, and depending on how these agencies decide to handle the matter, may lead to a more restricted program or to increased costs for families who do pay for services. Government reluctance to make defaulters pay may be a disguised form of subsidy to project participants. Although such a tactic may permit low-income families to live in a project who might otherwise not be able to afford it, the costs will eventually have to be met through reductions in government expenditures (including these programs) or with revenues collected from other sources. It is difficult to isolate any factor as

the primary cause of default: it involves affordability concerns, the adequacy of administrative machinery, and the role of political pressure groups.

Cost recovery in World Bank projects has varied greatly. In upgrading projects, the coordination between the installation of services and the beginning of payments is an important issue. In the projects evaluated, there have been intervals of up to two years between these two events. Obviously, beneficiaries are likely to be upset if they start receiving bills before there is water at the tap or working lights. What is not so obvious is that any substantial lag between commencement of services and that of billing and enforcement may seriously jeopardize cost recovery by giving rise to participants' expectations that they will never be made to pay. Precluding this potential dilemma is a major task for the project planner. Planning and implementation must be integrated, so that a decision can be made on when to start charging, given that the effective dates of service installations are uncertain and may be spread over several months.

Until implementation performance has risen to a certain standard, responsible agencies would be unwise to be aggressive in applying charges. Thus, the risk rises of engendering the assumption that charges will never start, or will never be enforced if they do. The chances of dealing effectively with the situation depends on participants' expectations, which depend in turn upon the city-wide context. What happens elsewhere when people default on service payments? In Zambia, for example, there has been a long tradition of

subsidized housing that began under British rule. There was clearly an element of income in kind in this type of subsidy. It has continued, and the subsidy is considerable at all levels of income and housing quality. Obviously, an effort to introduce a project seeking full cost recovery into such a context (especially with relatively poor participants) was going to come up against some pretty unfavorable expectations. The bulk of the cost recovery problems in Zambia are traceable to these matters of local tradition and expectation.

Cost recovery problems in Lusaka have been fairly severe. As many as 50 percent of families in some upgrading projects are in arrears, and some households appear to have paid not at all. Whilst affordability problems may be relevant in some cases, evaluation studies report that income was not a factor distinguishing defaulters from those who have paid, that in fact many defaulters had previously been paying considerably higher rents to landlords who could and would have evicted them if they had failed to pay.

A more direct factor in the slow rate of collections appears to be want of political will to seek payments (given the background of generalized subsidy in the housing sector) on the part of the Zambian United National Independence Party (UNIP). While the UNIP did help win acceptance for the projects, particularly in facilitating the potentially volatile transfer of families to overspill areas, it stopped short of effectively promoting the obligation of communities to repay project costs to the Lusaka City Council (LCC), the administrative agency. Though the LCC itself was somewhat remiss at the outset, it has

since taken steps to recover costs more systematically. A hindrance remaining is the lack of an efficient record-keeping system for up-to-date information on defaulters and on amounts owed, as well as lack of flexible mechanism for collecting payments. Families do not appear to be well informed about the payment schedule or about the purposes for which funds will be used. Notwithstanding all the above, the Bank's Lusaka project repayment record is considerably better than that of other projects administered by the LCC.

Because of the FSDVM's non-profit status, requiring it to effect full cost recovery in order to stay in operation, and because of its inculcation of social responsibility as a feature of project participation, the Foundation has achieved an excellent repayment record for its projects in El Salvador. Through mutual help and collaboration on important community matters, families are brought to understand that payments are a necessary part of the survival of project facilities, and how specific charges relate to specific services. This is particularly important where, as with street lighting, benefits vary with location, a distinction which was important to participants in Zambia, but had not been prominent to project planners. The FSDVM effectively uses a series of incentives and penalties to achieve cost recovery, including the support of community organizations to help collect payments, careful screening of participants for their ability to pay project fees, and the use of lawyers who visit families three months or more behind and pressure them to pay. Eviction is also an option open to the Foundation, though it is very seldom used. Technical factors, including

a computerized system of monitoring outstanding debts, and the fact that the FSDVM is small enough to supervise financial matters closely, have added to its success in gathering payments.

As a result of all these and other factors, the FSDVM has the best loan repayment record of any World Bank financed shelter project to date. As of July, 1980, total payments in arrears represented only 2.5 percent of the total loan portfolio. More significant, the great bulk of overdue accounts were only 2 or 3 payments in arrears. Only 22 percent of families in arrears were more than 90 days behind.

So, even though there are some non-replicable features of the FSDVM experience, it is plain that good design, effective implementation and maintenance, attention to community participation throughout, and general good management significantly enhance prospects for cost recovery. Further, these experiences suggest the wisdom of designing projects in a local public finance context, where the capacity to operate and maintain infrastructure and services can be directly related to collections or cost recovery, and a culturally appropriate mixture of incentives and sanctions can be sought that will bring payments into an affordable, collectible system. Collection mechanisms ought to be designed with community participation in mind from the outset, with regular, relatively convenient payment schedules explained thoroughly. Participants must be made clearly aware of what they are being asked to pay for, and for the consequences of default to themselves and to the project as a whole. Methods for improving cost recovery developed in

other Bank projects include discounts for prompt paymemy and the withholding of construction loans from families defaulting on lot installments.

Community Participation

As suggested above, community participation can play a large part in the efficient achievement of project objectives. While popular support and consultation may be sine qua non for eliminating resistance to project interventions, lengthy discussion of procedures or the training of project promoters in organizational skills can lengthen the implementation period substantially. Though project costs can sometimes be substantially reduced if community groups accept responsibility for the maintenance and operation of facilities, devising the administrative structure for such purposes can itself be costly.

Community participation should not be conceived as an indivisible, programmatic feature of project implementation. Some community institutions may not be representative of project populations as a whole, and those claiming to be so may in effect hinder accessibility of participants to project management. Similarly, whilst community leaders may be nominal or titular heads of social groups, they may not consult their membership on a regular basis and thus underemphasize or ignore minority factions or positions within their own groups. Because community structures vary greatly across cities and countries, organizations particularly effective in one setting may not be so in another.

What frequently happens is that intractability on the part of target populations, especially where resettlement is involved, prompts housing authorities to support the use of community organizations to help in dealing with the difficulty. If the effort succeeds, then they may look to the same organizations to carry out various other aspects of the shelter program, based on those organizations' knowledge of community needs and preferences. Bank projects have witnessed both positive and negative effects of community involvement (and lack of it) in project implementation. The FSDVM has compiled a noteworthy record of fostering popular responsibility for project facilities and services, yet there have also been cases in which groups have organized themselves against the FSDVM, refusing to make repayments until basic services were provided. An active community is thus not a passive tool of program management. Community organization can act either as a stimulant or a deterrent to reaching project objectives, depending on the degree of concordance between popular and program goals.

The Lusaka project has also used community involvement to advantage in some cases, and failed to employ it sufficiently in others. Consultation with groups in upgrading areas and with those scheduled to be moved to overspill areas avoided major disturbances in squatter areas traditionally considered areas of great unrest. Community groups were organized at the planning stage into committees which succeeded in making recommendations for the modification of road layouts that were acceptable to all parties. These committees have continued to be instrumental in other decision-making functions

affecting community well-being. Other components, including mutual help, have not been so successful. Though community funds were successfully raised to build and operate clinic facilities, other efforts, such as provision of multi-purpose community centers, have not proved to be popular. It is also probable that cost recovery would have fared better if communities had been consulted well in advance of attempts to collect charges, rather than after problems had already surfaced.

As for the Philippines, it appears that consultation with barangay leaders and community groups had led to the avoidance of major social disruptions during the reblocking process. It may also be that these Tondo organizations can be of use to project management in collecting service charges and payments, although the outcome is as yet uncertain. The evaluation team has noted a marked passivity in group meetings where project plans have been discussed. Tondo dwellers seem regularly to follow the barangay leaders' suggestions in declaring preferences for particular services and procedures. Although this form of hierarchical organization may in fact lead to acceptable results and the orderly recovery of costs from project families, the evaluation team has recommended that stronger information campaigns be conducted prior to and during relocation, so that households understand as completely as possible the options open to them. Such options may include actions that do not necessarily square with the preferences of the barangay leaders, but inclusion of such choices may lead to a broader-based community participation than the already estimable "grass-roots movement" already achieved. The most practical way of involving an

community in decision-making is to give its members the opportunity to choose among several options early in the project process. The Philippines experience to date provides a limited example of success in moving toward this idea.

Chapter 6:

SUMMARY OF PROGRESSIVE DEVELOPMENT EXPERIENCE AND RECOMMENDATIONS FOR FUTURE PROJECT DESIGN

It is now worth recapitulating the principles of the World Bank's approach to Third World urban housing needs. Progressive development remains the key concept. It is best defined as a method of increasing or upgrading housing stock that stresses: staged development; flexibility as to housing design, construction time, and materials used; and some form of self help. The self-help component has several possibilities of organization. It may take the form of mutual help, in which management administers work groups of householders. It may take the form the term "self help" implies, households building their own dwellings, perhaps with technical assistance from project management. "Self help" also applies to situations in which individual families hire contractors to build their houses, or hire and administer the work of individual laborers.

In the context of progressive development, self help may be understood as "self contracting." The chosen method of construction (where a mutual-help stage is not enforced) then becomes an efficiency issue, where choices are determined by a household's construction skills, opportunity costs, and access to funds or credit.

Generalizations from World Bank Experience

Casual observation and rigorous evaluation both confirm that potential consumers behave in a variety of ways in housing markets. The planning of construction, as well as the size, finish, and use of a

house may vary enormously, depending on legal, financial, cultural and other circumstances. This fact argues for market solutions to housing needs permitting households a maximum of flexibility. It strongly implies that project planners should not try to build everything into a project, but be selective, thinking of projects as interventions which "shock" the system to demand, and eventually produce, complementary services appropriate to locally defined needs.

On this view, then, projects should be kept simple, striving for realistic design objectives and speedy implementation. Key elements, such as appropriate locations and efficient credit mechanisms, must be included. The approach we are describing does not deny the complexity of urban problems, but rather interprets the complexity correctly. It requires careful and intricate analysis, relating objectives to resources and constraints. This process must lead to more accurate choices regarding what to do first in a given set of development circumstances. Coupled with an integrated handling of maintenance, cost recovery, and community participation, the progressive development approach makes possible rapid improvements in housing conditions for large populations, as well as replicable design and operations.

Our experience to date has shown that it is possible to produce housing affordable to at least some households in the lowest pentile of most urban income distributions, even though the projects supported by the World Bank have not always done so. ^{1/} When they have,

^{1/} The early Bank-supported projects, including those here evaluated,

the bulk of the project benefits have still not penetrated down that far. These appear to fall between the 20th, or perhaps the 30th, and the 60th percentiles in most cases. These figures will have to improve if we are to meet our equity as well as efficiency objectives rapidly and directly over the next decade or so.

There appear to be two principal ways to improve the general approach to make it more conducive to this end: one is to push standards, and therefore unit costs, still lower; the second is to increase the opportunities for rental income and occupancy. These two possibilities are not necessarily complementary, so they need to be considered cautiously and thoroughly. With regard to standards, there are two crucial issues: density of land settlement and service levels. ^{1/} The first matter is closely tied to the question of plot size, though the two concepts are not identical. Given the scarcity of urban land, in large metropolitan areas especially, projects should seek to achieve densities that are as high as possible, consistent with the economic demand of the target populations and reasonable health standards. This implies smaller plot sizes in future projects than in those already launched.

Smaller plot size does not imply reduced living space per household or per individual, as optimal plot sizes will be arrived at

^{1/} The timing of the introduction of various services is a third important consideration. In principle it should not be, for efficient credit markets would amortize the costs of the optimal solution as required. However, this problem should be kept in mind by project designers because credit markets are not in fact efficient.

through a number of considerations, including land price and the incremental costs of additional storeys. To date there has been an unfortunate bias in Bank supported projects against multi-storey construction. The reasons for this are not clear, but it may stem from the understanding of self help as self construction by householders, and from the fears and false hopes for this method. Early on, there was a tendency to believe, contrary to the evidence from many informal housing areas, that a significant number of families would "botch" a second storey.

Keeping the unit price of land within limits affordable to the target population must be a central objective of project design. El Salvador, where both the FSDVM program and the use of evaluation are of long standing, offers an example of the fact that a project's land provision and affordability components can be significantly improved through modifications.

The modifications have been of two kinds: improvements in layout increasing the total residential area; and stimulation of two-storey construction. The former derive from reducing the proportion of land assigned to vehicular traffic and parking by limiting these to peripheral areas, and the grouping of houses around mini-parks, green areas providing access to interior plots and semi-private recreation areas. These measures mean that families trade off some private area for semi-private area. The collective benefit is maximal in large projects where the mini-parks can be effectively complemented by larger open areas. The success of these policies is affirmed by the fact that

residential areas as a proportion of total area has increased from an average of about 50% in the earliest projects to 70% in the most recent ones.

At the same time, there has been a steady increase in the number of units per hectare of residential area. The greatest improvement in this regard has occurred in a first experimental project (called _____) in San Salvador, where two-storey design has made possible 140 units per hectare, as opposed to the 80 to 95 per hectare in other recent projects.

The other ways of reducing unit costs have to do with avoiding unnecessary construction costs and undesired services or service levels. The diversity of demand requires a similar diversity of options within projects, as well as flexibility with respect to their use. The extent of core house construction will depend on the relative costs of this versus other construction procedures. Precise estimations on these costs are still being prepared. Only a portion of plots should have any core construction, as our research results have shown that about half of all families choose to construct their own houses, at least in part. Still, it may be advisable to provide some construction on all plots for other reasons, such as to minimize reticulation costs, or to set up common walls to achieve particular housing layouts. ^{1/}

^{1/} Recent studies in the U.S. have indicated, for example, that housing costs are very low in Philadelphia. This has been explained by the high percentage of row houses in the city, and the efficiency of this type of design.

There are several related questions pertaining to service levels. In the case of water, the question is essentially whether to have individual household connections or some form of communal supply, such as standpipes. ^{1/} The problem is analgous to that of core house construction. Projects should first provide options. Not all households should be forced into a high cost option, as some are very likely to prefer some form of communal arrangement at a lesser cost. What proportions and what options are essentially empirical demand questions.

Among the most critical technical issues in urban development research today is that of finding more economical means of human waste disposal than water-borne sewerage systems. Until such are devised, the range of options is severely limited. Pit latrines, privies, etc., can be used in small cities and towns, and temporarily, in the peripheral areas of large cities. Some societies may be persuaded to attempt communal solutions on a limited basis. But generally the design task is one of keeping this expensive component within affordable limits. ^{2/}

Even with careful control of costs and the provision of low-cost options, a substantial proportion of families among the poorest will be "reached" only to the extent that rental accomodations are expanded along with ownership opportunities. This is because in any

^{1/} Of course, there are many other questions, such as capacity, hours of service, water quality, etc. However we are concerned here with project design.

^{2/} See, for example, John M. Kalbermatten, DeAnne S. Julius and Charles G. Gunnerson, "Appropriate Technology for Water Supply and Sanitation: A Summary of Technical and Economic Options," World Bank, December 1980.

society a significant minority of people prefer to rent. Evaluation research has not indicated any need to stimulate the rental market, with the exception that appropriate improvements in the credit system would probably increase the elasticity of supply of rental accommodations. The most important thing to do is ensure against avoidable constraints, such as the restrictions and prohibitions designed into some of the early projects.

Project designers must also bear in mind that there may be some conflict between the health objectives embodied in reasonable density standards and the rental objective. Any such conflict is unlikely to have serious consequences, however, as there is always the option of building upwards. In the event of actual problems, the range of plot size options can be extended at the high end, taking care to obtain full cost recovery for the larger lots.

The widely held assumption that rental solutions to urban housing problems are not adequate is misguided, given the evidence that large proportions of people in all countries choose to rent. The assumption is based on perceptions of squalid rental housing already existing in many less developed countries, and on the fear that new rental housing will further encourage exploitation of the renting poor by a rentier class. This phenomenon does exist to some extent everywhere, but it can be successfully combatted by expansion of the

total housing supply, ensuring that full cost recovery is enforced so that there is no tacit subsidy of a rentier class, and monitoring for specific abuses. ^{1/}

Finally, it will be necessary to experiment with and evaluate other means (than renting out housing space) of enabling the poorest households to use their houses and plots as earnings assets--means such as commercial or small manufacturing enterprises, or food production.

In seeking improved combinations of project components to serve the needs of target populations, flexibility will be the key. Preferred solutions will vary across countries because circumstances do, and even within a single city or country, optimal project designs will offer a range of possibilities.

The outlook for progress in project design is encouraging, given the demonstrated willingness and capacity of project planners (in the Bank and in national housing institutions) to interpret and apply the results of experience and evaluation. For example, almost immediately after adopting the sites and services concept, realizing its limitations, the Bank developed the complementary area upgrading approach. Recent years have also seen a significant broadening of options under both approaches. The modification of concepts has led to change from materials loans to construction loans, and acceptance of the need for rental components, as well as numerous other adaptations. El Salvador so far provides the best illustration of such adaptability.

^{1/} To guide this aspect of project activity, we need to analyze the indirect effects on the housing market of changes in supply wrought by the projects.

There the FSDVM has modified its mutual-help program significantly, pushed its already low standards still lower where necessary, and developed the layout and design innovations mentioned above. ^{1/}

Experience to date suggests that fresh consideration be given to the relative strengths of the sites and services and upgrading approaches. As Table 4.1 confirms, squatter area upgrading is not an unambiguously superior device for reaching the urban poor, as it was thought to be in the late 1970s. In the sub-projects studied so far (not numerous enough to be really representative) they have proved superior in reaching the poor, but not in excluding the more affluent, which is one of their corollary objectives. Thus the preferred approach will depend upon how one weights the benefits distributed to various income strata, and on the scope for improvement in sites and services targeting.

We have seen that adopting a total income definition and adapting project design to make lower cost options available will allow inclusion in projects of significantly larger proportions of the urban poor, without compromising affordability objectives. If the expected demand response occurs, then income ceilings can be lowered without curtailing the scale of projects. Though the results are not assured, this is a promising area for experimentation and evaluation.

^{1/} The other project countries provide fewer specific examples because there have so far been no "repeater" projects in Zambia and Senegal, and the Philippines project is of much more recent origin than the rest. Still, a similar pragmatism has been evinced in responses to implementation problems.

The political dimension of such a trade-off must also be considered. Most politicians are less reluctant to accept low standards in upgrading projects (where almost any change constitutes an improvement) than in sites and services developments. This bias can further complicate the already sensitive problems of acquiring land suitable for sites and services schemes. There is currently less resistance to upgrading proposals, and institutions such as the World Bank are prone to follow paths of lesser resistance. This is a tendency that must be opposed, and fortunately experience and evaluation suggest a way.

Area upgrading and sites and services are really part of a continuum of nuanced possibilities, not really distinct approaches at all. The processes involved in reorienting a wooden house on a Tondo lot or moving it to an overspill area, or even to the Dagat-Dagatan sites and services area are really the same. To be sure, time, distance, cost, and the degree to which the original structure is retained will all vary. The point is that they vary along a continuum, frequently with very small differences. There are extremes, in the form of very large new developments with substantial house construction, for example. However, it is arguable that such activities have a place even in programs of marginal housing improvement. Leaving such extremes aside, planners' efforts should be to get politicians to focus on the

continuum, and on the aim of improving the total housing stock gradually over time. 1/

Maintenance, collections, and cost recovery--and their inter-relationships--must be vital considerations for design and management. These matters must be dealt with from the outset, and projects should not be designed that do not fit comfortably within a general public finance framework, in terms of the value of benefits realized relative to the revenues extracted. Projects must be designed so that they return more to the public revenues than they require as expenditures, including maintenance and operating costs. This is the essence of the replicability issue.

Because of the difficulty of this objective, further experimentation with community participation is a must. Fortunately, the record on this score is quite favorable and encouraging. Just as they have the desire and capacity to improve their own housing, the urban poor have a desire and capacity for greater participation in their own governance. The World Bank projects have shown that such latent capacities can be released. It will be the task of future project designs to carry his promising process further.

Research Issues

All the issues discussed here call for continued, selective evaluation. The Bank's formal research program will concentrate on

1/ To developing politicians who are suspicious of "capitalist foot-dragging" on social objectives, a bit of the Soviet experience in dealing with housing problems may prove instructive. [Ref. H. Smith, Ch. 3].

issues of demand for housing and related services over the next two years. This work will proceed on three fronts: general studies of demand; studies of the attributes of housing; and studies of tenure choice and the demand for credit. The general studies of demand will be completed fairly quickly. The objective will be to extend the estimates of the central demand parameters (income elasticity and average propensity to consume) from the current base---Colombia, El Salvador, and Korea---to a broader selection of countries and cities. This is necessary to provide reasonably accurate affordability benchmarks for project design in the full range of circumstances within which the Bank is and will be operating.

The need for the two other research undertakings is obvious in guiding a program based on simple sites and services and compatible credit schemes. The studies of housing attributes will use hedonic analyses to estimate the values of participant attributes to various aspects of housing, in varying circumstances and combinations.

The studies of tenure choice and credit demand will attempt to define the bases for developing more effective credit instruments, including programs to foster a more active rental market. They will be coupled with experimentation, evaluation, and analyses on the supply side.

PART III

Chapter 7:

ESTABLISHMENT OF THE DEDRB EVALUATION PROGRAM

This and the following chapter describe the experimental evaluation program applied over a five year period to several of the urban development projects financed by the World Bank. ^{1/} Although it shared features with other sub programs attempted elsewhere, three unique aspects serve to distinguish it from other, comparable efforts. First, the program has been executed by a research division of the Bank, collaborating closely with its counterpart projects department. ^{2/} Second, it has had its own financing (rather than beign financed with project funds), which led to collaborative arrangements with the International Development Research Center (IDRC) of Canada. Finally, it overdid a good thing in the sense of being somewhat premature, relative to project design and implementation experience.

^{1/} Up to 1972, evaluation activities were referred to as "monitoring" within the Bank departments involved. After April, 1973, the terms began to be used in conjunction, or interchangeably. Since April, 1976, only the term "evaluation" has been used by those involved in this program, on the assumption that all effective evaluation systems will embrace monitoring.

^{2/} A few words on bureaucratic organization and titles are in order here. In October, 1972, World Bank management established a Central Projects Staff, and renamed its Central Economics Staff the Development Policy Staff. The Urban Projects Department has always been in the former. However, it has had several names: Special Projects Department, until October, 1972; Urban Projects Department Until October, 1973; Transportation and Urban Projects Department until March 1976; again Urban Projects Department until July 1981; and, since then, Urban Development Department.

Origins of the Program

As the Bank's first "urban" loan, for a sites and services project in Senegal, was in final stages of preparation, the Urban Projects Department and the Urban Research Division shared a growing concern for verifying the economic effectiveness of this type of project. Though both assumed that projects like the one in Senegal would enhance the incomes and welfare of participants, they also knew that research into the question had been very limited. The appraisal report was guarded and general in its claims. ^{1/}

The early arguments for monitoring and evaluation hinged on the idea that such research efforts would yield ways of measuring the effects of project participation on households' real incomes. This issue was taken up again in the Board discussion of the Senegal project in June, 1972, in which consideration was also given to whether ways of measuring project impacts on health, education and employment could be developed through the same process. A follow-up study was requested, and the Projects Department and Research Division jointly produced a first design of it during the summer of 1972. Interestingly, the statement of objectives formulated then remains valid. ^{2/}

The general development of an evaluation strategy for urban projects succeeded in most respects conjointly with the devising of a specific research program for the Senegal project. The process of

^{1/} Appraisal of a Site and Services Project, Senegal, Report No. PS-11a, Special Projects Department, June 7, 1972, pp. 22-26.

^{2/} Appendix Ii recapitulates the statements of objectives put forward during the first years of this venture.

research design began with the first project supervision mission to Dakar in October, 1972. After visits to research institutes in Dakar and discussions with researchers there, tentative recommendations were made as to the scope of the projected study. The most important conclusion drawn at that time was that a system would have to be devised to permit the flow of information and analysis between various interested parties: the project agency, an eventual field study team, and the Bank.

The continued involvement of the Research Division in project supervision further promoted the view that it was essential to answer, or at least to address, as wide a range of questions as would be feasible. It was also agreed that field studies of only one project would not constitute an adequately representative base for evaluation of urban shelter development. A proposal for a more fully elaborated research design was presented to the vice-presidents of the Central Projects Staff (CPS) and the Development Policy Staff (DPS). It called for monitoring and evaluation of three projects to be supported by the Bank, once suitable projects could be chosen according to the timing of their execution. This proposal was accepted in January, 1973.

Nearly three years passed before an additional two suitable projects had been specified, a specific financing commitment had been made, and work programs agreeable to all parties had been formulated. The Zambia project was approved in June and the El Salvador project in September of 1974; the financing was committed in January, 1975.

Meanwhile, despite substantial efforts, the Senegal project languished. Co-financing arrangements had been worked out by the Bank and IDRC. And substantial progress had been made in the areas of research design, personnel, institutional development and affiliation, and operational choices concerning how the work was to be managed from within the Bank, and how project execution, supervision, and evaluation activities were to be related.

Starting from the DPS-CPS agreement about initial study design, the Research Division made a broad range of contacts with institutions and individuals familiar with evaluation methodology and urban research. The revised preliminary research design was sent in early 1973 to some 35 specialists from Africa, Europe, and North America, each of whom had some of the expertise necessary to conduct a successful evaluation. All were asked to respond if they were interested in being involved in such a program, and to suggest others who might be qualified. There was quite a reaction, but qualitatively it was disappointing. It soon became clear that very few researchers combined the requisite methodological skills with necessary fieldwork experience in urban Africa.

At the same time, the complexities of a sites and services project--including housing, public utilities, social services, and financial arrangements for a project population of 140,000--required expertise in the evaluation of shelter and community development schemes that initially did not appear to be available within the Bank. Discussions were held with several outside groups and eventual agreement

was reached with an interdisciplinary team of scholars from eastern American universities, individuals who had dealt with various facets of evaluation in other contexts. This team agreed to develop a preliminary design for monitoring and evaluating urban projects.

In spite of their wide experience, the consultants were unable to improve substantially upon the framework for evaluation developed by the Research Division in 1972 and 1973. Though the consultants were required to produce a preliminary statement in only two months, the time constraint did not appear to be the major factor in their disappointing performance. Much more important was the fact that even such excellent team of consultants proved unable to inform themselves about, or keep abreast of, the Bank's burgeoning interest in sites and services projects, despite extensive briefings and sharing of information.

This assessment of the consultants' statement eventually led to the decision to dispense with them for the most part, and have the evaluation work done "in house." Rather than have consultants design and execute the work under staff supervision it was decided that staff should perform the work, drawing upon the specializations and advice of consultants only when necessary.

This decision reflected the view shared by the DPS and CPS that the Bank already had a comparative advantage, thanks to the experience and caliber of its staff, over consultants, even if the latter were among the best in their field. In retrospect, this decision proved its wisdom. However, it is interesting and instructive to consider how different were the interpretations of the decision at the

time. Bank staff took it as a signal to roll up their sleeves and deal seriously with the obstacles to research. The consultants took a different view, regarding the decision as the Bank's effort to develop an evaluation program "on the cheap," without the aid of appropriately qualified social scientists. ^{1/}

It was necessary for Bank staff to maintain supervisory responsibility, particularly in determining the content of fieldwork and its procedures, in order to address the issues surrounding project design and the conditions affecting urban lending policy. It was equally important to develop an appropriate institutional and training relationship with individual project contexts, something consultants could not be relied upon to do.

The proper "administrative distance" between evaluation and project implementation was recognized as a crucial issue early on, and continues to be a concern. Bank staff were insistent that difficult and sensitive analysis be carried out according to clearly defined methods, yet they put out ambiguous signals, so as not to discourage personnel and researchers from using the local initiative necessary to analyze and resolve strictly local problems no research design could foresee.

Research Methodology

The design of monitoring and evaluation field studies underwent considerable evolution between the initial work of 1972 and the establishment of field teams between July, 1975 and January, 1976. Throughout many discussions, and the preparation of a series of papers

^{1/} Critical abstract from correspondence.

on the subject, there was a continual tension between comprehensiveness and concentration. The variety and scale of project components could easily justify substantial study efforts for each component. However, given the constraints on resources and the need to identify the most important questions in this effort, certain concessions had to be made. Researchers successfully held out for certain procedures that permitted a reasonably systematic and rigorous approach. These included lengthening the period of analysis to a minimum of five years, and the extension of the study design to include appropriate samples of control households in other neighborhoods and housing environments.

The initial conception of fieldwork incorporated baseline surveys of the socio-economic characteristics of project applicants; sample surveys at the plot/household level to determine changes in plot and household characteristics; and special studies of project components, such as house consolidation (lot development) or employment creation, to answer questions not covered by the household surveys. Changes within the project area and its population would then be compared with surveys of control areas and groups within each city. Bank-supported field teams would stay in the field for at least two years and train local counterparts to carry out the work over a longer period. Sample surveys would be carried out at regular six-month intervals in order to capture all phases of housing consolidation and economic change at the household level. The initial two-year phase of Bank-supported work was to be supplemented by research projects and small studies generated in the course of the work. Central to this

approach would be an affiliation with a local research institution that would store data, train interviewers, and eventually make the effort a permanent local program.

Though this general research framework was acceptable to many people consulted on the matter within and outside the Bank, it became clear to everyone upon reflection that many significant behavioral changes in the project area would not occur within the first two years of project execution. Further, implementation delays could mean it would be several years before "reasonably complete" structures were built on the majority of individual sites. This was thought likely to be the case in Senegal. ^{1/}

After discussion with researchers who had attempted surveys of this magnitude in similar conditions, it became clear how difficult it would be to achieve the short turnaround time required to carry out surveys at six-month intervals. Many advantages were seen in lengthening these intervals to periods more consistent with local conditions and which could yield a more informative view of the housing consolidation process. It was therefore decided to conduct surveys at years 1, 3, and 5 of project execution. The first would be a baseline survey, linked with project participation selection, to avoid duplication of effort. The decision to extend the length of Bank-supported field work also created opportunities to broaden the scope of analysis to include such elements as family employment and income

^{1/} Appraisal of a Site and Services Project (Senegal), Op. Cit., p.ii and p.14.

profiles in which changes might not register until long after a household joined a project.

There was much discussion of the use of control groups at this time also, and two important changes of approach resulted. The first was related to the extended length and scope of the analysis. The lengthened period of research meant that control groups would have to be chosen very carefully to maintain their validity over five years, relative to a wide range of variables. Though it has not yet prove feasible, relating both controls and project populations to a sample frame (the city population as a whole) was deemed a desirable goal.

Secondly, as squatter area upgrading was adopted as an instrument in the Bank's urban shelter strategy, researchers realized that comparisons could be made as easily between upgrading and sites and services areas as between either of them and outside controls. Both the magnitude and rates of changes in behavior could be compared to determine whether new elements in new project environments or new elements in old, familiar environments produced more favorable outcomes. Evaluation of the two shelter strategies could then be combined in a single effort, with only some modifications of the survey instruments to be used. (In practice these arrangements worked out somewhat differently than was foreseen.)

All the refinements of the initial research design were intended to respond to fundamental methodological issues of inference and statistical measurement. The Urban Research Division kept abreast of the most recent advances in evaluation studies and field research

methods. Its staff and consultants produced a series of notes and paper abstracting and summarizing this knowledge. This phase of work culminated in 1976 with the production of a long and comprehensive set of methodological guidelines. 1/ The guidelines pointed out a broad range of methodological hazards that could seriously compromise the validity of results. The paper was subsequently circulated among those involved in the technical design of evaluations, both in the Bank and in each field setting. Thus began the Research Divisions ongoing role as "methodological guru" for the entire exercise. 2/

One of the most important aspects of research design worked out between 1973 and 1975 was the specification of independent, dependent, and intervening variables for the analysis. This process was a first step in the effort to distinguish effects on project populations attributable to project components from the effects of population characteristics and behavior on the "success" of project components. 3/ In evaluation parlance, the problem is one of reckoning internal validity, external validity, and relative effectiveness. 4/ Discussion focused on project components' characteristics as independent variables, magnitudes and rates of change as dependent variables, and the socio-

1/ This effort drew heavily on Thomas Cook and Donald Campbell, "The Design and Conduct of Quasi-Experimental and Free Experiments (six) in Field Settings," in N.D. Dunette, (ed.), Handbook of Industrial and Organizational Research, Chicago, Rand McNally, 1975.

2/ Michael Bamberger, The Primary Evaluator for the El Salvador program, in 1975-78 and currently a World Bank staff member, oversaw much of the methodological refinement of the evaluation program.

3/ The question of this distinction was first raised in 1975 by Professor Remi Clignet.

4/ See the Glossory appended to the text.

economic characteristics of the population as intervening variables.

The early lists of these variables are presented below:

A. Independent Variables

Provision of plots.
 Characteristics of plots (plot size, plot options)
 Form(s) and availability of credit
 Incentives for self- and mutual-help
 Provision of community facilities (e.g., schools and markets)
 Service standards for project as a whole
 Technical assistance components.

B. Intervening Variables

Socio-economic characteristics of populations (e.g., age, sex,
 education level, original income)
 Migratory and familial history of populations
 Characteristics of the city and the urban environment

C. Dependent Variables

Housing investment
 Rate and nature of consolidation of plots
 Changes in household budgetary patterns, including income,
 consumption patterns and savings
 Changes in access to services
 Rates and reasons for defaults and attrition

This specification of variables preceded formulation of preliminary hypotheses. Field teams were charged with formulating detailed sets of hypotheses to be tested in individual research designs and field settings. These would vary with the characteristics of project components. The interim research designs developed during the spring of 1976 are summarized later in this chapter, following a description of the circumstances, institutional and financial arrangements, and the substantive guidelines for evaluations.

Institutional and Financial Arrangements

Monitoring and evaluation efforts in each country were to be affiliated with a local research center. This affiliation was intended to permit greater continuity, local contacts, training, and was expected ultimately to contribute to the development of the research centers themselves. Extensive visits were made to existing research institutions in Senegal, Zambia, and El Salvador to determine whether a favorable working context could be found there. In Senegal, this effort included three years of discussion with an inter-university research project run by the Universite de Dakar and the University of Leiden about the possibility of establishing an urban research center at the Institut Fondamental d'Afrique Noire (IFAN) at the Universite de Dakar. Similar efforts were made in Zambia and El Salvador to develop an institutional connection given the field team a sound research base unencumbered by constraints that might keep evaluation efforts from proceeding expeditiously.

In each country there was a tradeoff involving the administrative and physical distance separating evaluation from the executing agency. Too close a connection could result in pressures to produce a favorable evaluation, where as a looser connection might permit more objectivity. Then again, reduced distance could also mean better access to information and a clearer understanding of the operational problems faced during execution. Previous experience in El Salvador indicated that "excess distance" could be crippling indeed.

As institutional affiliations in host countries were under study, negotiations were also underway with the International Development Research Center (IDRC) to secure its participation in the funding of the evaluation program. At the suggestion of the IDRC, a conference was held in mid-July, 1975, in San Salvador, El Salvador to convene the directors of the executing agencies, the primary researchers, IDRC staff, and the appropriate operations and research staff of the Bank. The conference had two major objectives: to explore the nature and methodology of the proposed research, and to move towards decisions on institutional and financial arrangements within each country and between national and donor agencies.

With regard to the first objective, there was considerable progress during the conference. It was decided that a new office of evaluation would be instituted in the FSDVM, the executing agency in El Salvador, and that the research team including the primary researcher would be directly under contract to the Fundacion. Owing to previous evaluation the experience mentioned above, which produced only a final report, the Fundacion's directors felt this arrangement was necessary to guarantee that the evaluation in this case would yield frequent reports that could be of operational value.

For Senegal also, it was decided to locate a Bureau d'Evaluation within the executing agency, the Office d'Habitat des Loyers Moderes (OHLM). The primary researcher would direct the Bureau and report to the Director-General of OHLM. The Bureau would be staffed

by an associate researcher and interviewers, permanent employees of OHLM, and financed by the Bank and IDRC.

In Zambia, the evaluation team was to be established within the Ministry of Local Government and Housing to which it would report directly. However, the team was in fact physically located within the executing agency, the Housing Project Unit, where it was to cooperate actively with the HPU's own monitoring unit.

The formation of steering committees was strongly recommended by the IDRC to deal with the threat to objectivity arising from close connection between evaluation and implementation staff, and to facilitate dissemination of evaluation reports. The recommendation was readily accepted by the Bank and the executing agencies. The steering committees were to operate somewhat differently in each country, but they were to be advisory bodies composed of individuals from universities and other research centers, various interested government agencies, and the two donors. The committees were to comment on the research designs and provide the evaluation teams with contacts to individuals and organizations that might assist the research effort. Actual direction of the research was carried on jointly by the Bank and executing agencies, and very few differences were noted in their

respective considerations of the questions the evaluation should pursue. 1/

Funding for the field research was to be provided in roughly equal shares by the Bank and the IDRC. The Bank initially committed \$450,000 to be disbursed over five years: July, 1975 - June, 1980. The IDRC initially allocated \$320,000 for the three-year period: October, 1975 - September, 1978, with a commitment to continue funding at approximately the same annual rate for another two years. The Research Division's supervisory role entailed additional funds for travel, consulting services, and computer time.

With project implementation expected to take years, it was assumed steady supervision on the part of the Urban Research Division would be necessary. Funds were provided from the budgets of the IDRC and the Research Division to bring researchers and IDRC and Bank staff together periodically -- approximately annually -- to discuss the course of research. The IDRC and the Bank agreed to share approximately in the financing of these conferences.

Guidelines

In consultation with projects staff, the Research Division prepared a comprehensive set of guidelines for the evaluation of urban shelter projects. This general statement of the Bank staff's view of

1/ The steering committees functioned more or less as planned for a time, and then ceased to function regularly. The smooth relations among groups involved in the evaluations resulted from the Bank and IDRC's continual participation in periodic supervision and in the annual conference. This institutional oversight has done more than enough to keep the evaluations focused and to maintain the objectivity of analyses.

evaluation objectives was circulated among field researchers in the fall of 1975 to guide the ongoing formulation of individual research designs. In general, two broad levels of analysis were planned for: a determination of whether overall project objectives had been achieved, and a determination of the efficiency and relative contribution of particular components.

At that time, the stated goals of the Bank's housing project sponsorship were to improve the physical, social and economic environments of the urban poor presently residing in squatter settlements and central urban slums. ^{1/} Intentionally operating with minimal government subsidies, the Bank affiliated agencies relied upon the following resources to effect these improvements: provision of tenure, basic infrastructure, technical and financial assistance, as well as health clinics, schools, and incentives to stimulate small-scale employment opportunities. Behind the commitment of these resources lay the assumptions that the institutions wished to test:

1. These resources would foster substantial inputs of self-help labor and would rapidly raise physical standards of housing above those that would otherwise be available; this would be accomplished at prices residents could afford.

2. Over time, project residents would enjoy net economic gains through increases in property values, increased opportunities for rental of rooms, improved employment opportunities through sites for small-

^{1/} Just as individual project and evaluation goals, these overall goals have continued to evolve.

scale industry and acquisition of skills from self-help efforts, with little or no adverse effects from residence relocation.

3. General levels of health would increase.

4. School attendance would increase and general educational levels improve.

5. Through active involvement in project execution and decision-making, residents would develop a stronger sense of community and consequent sense of responsibility for development and maintenance of community programs.

6. Through experience gained from these projects (with their minimum of subsidy), the cities and countries involved would be able to replicate the projects on a broader scale with indigenous resources, financial and technical.

Bank staff also wanted research to move beyond these considerations to gauge the overall impacts of projects on the cities and residents involved. What changes do households face in their patterns of saving expenditure? If expenditures for housing are increasing, what expenditures if any are decreasing, relatively or absolutely? What are participants' attitudes towards these changes? What changes are there in residential stability? Which families are moving, which are staying, and why? Are there any changes in family composition or structure?

With regard to projects' impacts on their host cities, it was deemed important to ascertain how such low-cost housing schemes could be fitted into planning for each city's future. Do they have an impact on

a city's spatial growth? What are a project's impacts on neighborhoods not directly involved in them? Do such factors as rising land values suggest that projects will work well only in particular phases of a city's development? At what point do other possible housing schemes become more attractive?

Though answers to these questions would come only after considerable time in the field, two relevant sets of data called for analysis in the early stages. One pertains to land values in and around a project, and participants' responses to changes in values. The other consists of records on both selected and rejected candidates for project participation, and any existing profiles of the city's poor as a whole. It would be necessary to know whether the housing and services provided were actually reaching the intended target groups in order better to estimate unfilled demands for low-cost housing within a city's low-income population.

Components

Beyond questions about general project objectives, it was important to understand the workings of individual project components. For each component, two basic questions had to be raised:

1. How efficiently does a component function in providing the specific services it was designed to deliver? As long as feedback was rapid enough, this line of inquiry had the greatest potential for aiding the ongoing work of project units and evaluation teams.

2. Irrespective of a given component's efficiency, how much does it contribute to the realization of basic objectives? What is the marginal

impact of the provision of one level of services compared with other levels, where this can be ascertained? The capacity to answer these questions would depend of course upon the different kinds of treatments or components within a given project. Without such possibilities of comparisons, it would be difficult to isolate the impacts of any particular component. 1/

Listed below are the specific components then included in Bank-financed housing projects, along with the questions, keyed to the issues of efficiency and project objectives, considered most important at the design stage. This list was not intended to be exhaustive, and field teams were expected to adjust the list to the concerns of their particular projects and agencies.

1. Infrastructures: How well are services maintained, especially in cases where residents have some responsibility for their upkeep? How quickly and to what extent do residents take advantage of optional connections for sewerage and electricity when these are available? Do residents find these services worth their investment? Understanding their need to pay for further services, what priorities do they have for additional infrastructure development? Must design standards be raised to produce desired impacts, or could they actually be lowered?

1/ In fact, there was little scope for such analyses within the first generation of projects, given the tendency to settle for homogenous packages of components, for the sake of administrative ease. Despite these difficulties, high priority was attached to the questions mentioned here because answers to them could be helpful in determining the appropriate allocation of scarce resources within projects.

The supply of fresh water, sewerage, and drainage combine to form a minimum public health package. The assumption is that all three are needed if appreciable health benefits are to result. The costs of these services are already known. However, there is still much to be learned about the levels and combinations of inputs required to produce significant results.

2. Components relating to the individual unit: The rate of housing consolidation was to be followed closely. What do project residents build, how quickly, and with what combinations of materials and labor options?

Affecting this process were several other factors, such as lease/tenure regulations. Are they clearly understood by participants? Are restrictions (such as preclusion of plot sale for a number of years) accepted, or is there resistance or call for changes? Is there a demand for rental arrangements?

Other questions relate to plot size and location: how is the plot space used, and how much of it goes to the house itself? Does location near a road or community facility affect the use or value of a plot?

The level of core construction and the building materials loan program are other factors bearing on housing consolidation. Are there sufficient materials of the right variety to supply residents' needs? Is the credit system understood? Do residents tend to borrow heavily or little? Do they stay within their means? Would the community's own informal supply networks function as effectively?

Technical assistance and labor use too affect housebuilding. Is such assistance sufficient and readily available? What are the impacts of specific regulations and guidelines? How much of their own or their families' labor do owners use? How much do they contract out? In the overview of housing construction and infrastructure installation, does it appear that some of what is assigned to self help should not be, or that some work assigned to contractors should be left to self- or mutual-help?

3. Financial arrangements: Do residents keep up with mortgage payments? If default occurs, is it due primarily to lack of income, dissatisfaction with results, or other factors? Is there a demand for a different repayment schedule? How are project services supported by residents paid for? Is there community acceptance or resentment of individual and/or communal collection systems?

4. Project location: What are the effects of relocation on employment among sites and services participants? Who shifts jobs and why? How much of a family's income goes for transport? Do residents generally have easy access to the city for services and employment?

5. Employment: How many residents take advantage of the small business, commercial development, and loan programs within projects? Are the enterprises these programs are meant to support viable? Is much local enterprise guaranteed? How much employment do nearby industries provide project residents? Do they generate support industries or services that provide jobs to residents? Does the project itself attract business to locate in its vicinity?

6. Community facilities: Are residents aware of health clinics and the services they provide? Does the demand for medical services generally exceed the supply? Are the schools adequate to serve the community? Do children attend? If not, why not? Do parents, teachers and the Ministry have goals in common? Has the provision of community facilities helped in the realization of community goals? Are the markets well located? Do they aid residents desiring to make small commercial investments? Do recreation and community centers fulfill their functions? Are they used?

7. Project layout: Do residents have adequate access to facilities? Which parks and open spaces receive the greatest use?

Evaluation personnel realized that the research analyses would have to control for the various social and economic variables (intervening variables) describing the residents at the time they entered the project. Age, status, income, employment, number of children, and certain other basic variables might explain more about why and how particular objectives were achieved than they would indicate the results of providing particular components.

Given the large number of questions of interest, it was clear that great care and economy would have to be exercised in collecting data and formulating research designs. It was deemed advisable to separate data requirements conceptually from the analyses carried out. The impacts of relocation on employment, household expenditures, and other items could be complex and should, of course, be explored fully

through crosstabs, correlations, etc. However, data needed on transport might well be small, e.g., just a few questions on distance and time and money costs of getting to work.

Considerable "compacting" might be accomplished since, as in understanding the process of housing consolidation, the information collected to test a general objective might be the same as that needed to evaluate a particular component. At other times, of course, this would not hold true. A utilization survey might be necessary to determine the efficiency of a health clinic, but other data, such as incidence of child mortality over time, would be necessary to determine whether the general objective of improving health care has been achieved.

Neither the list of components, nor the questions suggested were meant to constitute a straghtjacket on the testing of hypotheses. In the long run, for example, it might be found that health care may have a greater impact on incomes than any of the particular components dealig with employment opportunities. The rate of housing consolidation could be more significantly influenced by employment patterns and opportunities than by lot dimensions, locations or service levels. In short, it should be expected that interrelationships among components and between specific components and general objectives would be very complex.

Research Designs

Between July, 1975 and January, 1976, the preliminary research designs presented in San Salvador were revised in line with priorities

articulated by each country and executing agency, and with the guidelines just outlined that had been articulated by Bank staff. The designs were assessed critically during a technical conference held in Washington in January, 1976, and then revised and extended yet again over the next few months to yield working versions to be used during the first two years of project evaluation. It was felt at that time (and it has so transpired) that the designs should be as flexible as possible. Both the projects and the means of evaluating them were largely new and untested. Unexpected results would be plentiful. Thus there was a strong case for facilitating rapid adjustment of the study programs in response to changing conditions and research demands.

This section provides brief descriptions of the evaluation programs and their research designs, as viewed at the outset of the exercise. It is important to bear in mind the extreme diversity among the four projects discussed, in individual project characteristics, institutional arrangements and country contexts. ^{1/} The projects will be treated in the order in which their research designs were developed. The Philippines project is included here because, even though the project itself was not approved until May, 1976, preparations for it were made much earlier, and the program had in all respects "caught up" by the middle of 1978.

^{1/} The important "unifying factor" among the initial three projects was their location in countries whose national per capita incomes were very nearly equal in 1975, at approximately US \$400 per annum. The figure for the Philippines was somewhat lower.

I. El Salvador

Dimensions:

Of the projects chosen for study, that in El Salvador involved construction of the smallest number of units, but represented the most intensive intervention for those participating in it. Each plot would have connections for water, water-borne sewerage, and drainage, and virtually all would have a core element of a house built by contracted labor. Under the supervision of the Fundacion, the project reflected a concern for broader social issues. Mutual help, for example, was stressed as a means to strengthen family and community ties, as well as a procedure for lowering costs.

The establishment of an evaluation effort in El Salvador benefitted greatly from the prior existence of research programs and on-going record keeping within the Fundacion, as well as from studies previously conducted by government agencies.

The research design reflected this background. The proposed survey instruments were designed in part to query the community on preferences for different levels of core construction and on the problems and successes involved in mutual-help labor. Consistent with the Fundacion's concern for measuring social change, the evaluation unit intended to make use of less structured research methods (case studies, written observations of the social workers, etc.), regarded as better suited than formal questionnaires for developing a clear understanding of community values and attitudes.

The research design was structured along four main lines: the assessment of changes in communities; the assessment of the populations' and settlements' effects upon the project; estimation of the project's indirect effects on the city and other low-income settlements, and on local and national housing policies; and evaluation of the executing agency in the form of feedback to assist management in policy development. Certain specific research instruments were developed for each type of inquiry, but in all cases secondary sources of information were to be used along with freshly collected data.

Data and Evaluation Activities:

Direct observation, in depth interviews, and case studies would complement information obtained through survey questionnaires. Among secondary data, the most important were: 1) an evaluation study carried out by a local university, completed in 1975; 2) a housing market study undertaken by the Fundacion; 3) an employment survey conducted by the national planning agency; 4) an on-going OAS-financed study of building materials and technology; and 5) reports from various government agencies. The fundacion already had an extensive and relatively efficient management information system which was placed at the disposal of the evaluation. Data were being collected on a routing basis on project applicants, mutual help projects, beneficiaries of building materials, loans, mortgage collection and default rates, reports from social workers on community development and cooperatives, and legal data on withdrawals or evictions from the project. Thus, besides using information gathered directly by the research team, the

evaluation would rely in large measure upon data already available which, when systematically processed, would permit considerable expansion of the scope of analysis at low incremental cost.

Results:

Research results were to be communicated to the executing agency, the Bank and the IDRC through a variety of periodic reports. These would include: 1) brief monthly progress reports; 2) quarterly technical progress reports; 3) three more more partial evaluation reports annually, covering special topics such as the selection process, building materials, loans, mutual help, etc.; 4) methodological reports scheduled as needed; 5) full annual reports covering all dimensions of the evaluation; and 6) operational reports designed to answer specific requests made by the executing agency. In addition to these reports, the research team would participate in periodic technical conferences organized by the Bank and IDRC and would make available to Bank staff the raw data collected in the sample and panel surveys.

Status:

There were some clear differences between the research interests of the Fundacion and the Bank. Differences then were seen as a matter of emphasis, with Bank staff interested in indicators of housing quality, economic and employment-related issues, and the Fundacion putting more stress on social issues. The evaluation team did not anticipate problems in resolving these differences. In fact the differences turned out to be fundamental and substantial, requiring periodic discussions of evaluation goals over the ensuing three years.

The evaluation team did agree to collect more information than had been planned on areas of social concern, such as the degrees of financial stress families experience at times of relocation and initial housing consolidation. Beyond this kind of ambiguity in research coordination, only marginal adjustments in the research design were necessary. For example, Bank staff suggested that the sample for the experimental group be increased and the number of controls decreased.

The baseline questionnaire for the first community to be studied, Santa Ana, was already being applied as the research design was finalized, and it was expected that a report on the data generated would be available in the spring, and more fully in time for use by the Fundacion and the Bank in planning a second project. Preliminary results of the Fundacion's own evaluative work, as well as project supervisory activity, had already led to agreement that the second project would incorporate elements of increased "social experimentation," including alternate credit instruments and a broader range of service levels.

II. Senegal

Dimensions:

The major characteristic distinguishing the Senegal sites and services project from other Bank urban projects was the establishment on 400 hectares of some 14,000 sites and services plots, located 7 km from the city center. This location, and the consequence that participants would have to move from other residential neighborhoods in the city

would create a range of problems concerning transport that would influence families' participation in the project. The given conditions of the project presented that interested population with a set of logistical problems that would influence their capacity to build on the sites and meet their own needs. These conditions would be reflected in household composition, and in patterns of development and repayment of monthly charges.

Data and Evaluation Activities:

To understand the influence of project conditions on the processes of social adjustment and housing consolidation, a large range of instruments and data collection methods were incorporated in the research design. It composed 32 data collection operations to be carried out at different points during project execution. Within the various data collection operations, the field team would carry out sample surveys, in-depth interviews, and case studies. These would be based initially on the results of baseline census of all applicants to the project. ^{1/} In addition, the data collection operations would make use of studies previously carried out by other institutions, including the studies of Project RUL-12 and sample surveys by the Section d'Assistance de Communautés (SAC). Over the preceding decade, numerous urban studies had been undertaken in the squatter settlements of Dakar,

^{1/} [Explain why this won't work, and the need to rely on central census and statistical agencies.]

and these would be drawn upon as much as possible to support the evaluation effort. ^{1/}

In addition to the above, separate efforts would be made to understand the formation of community associations as this occurred within the project. These associations were expected to affiliate project participants into groups that would become the essential neighborhood units within the physical layout. Their cohesion and ability to work together, encouraging residents to improve the neighborhoods, was to be a major factor in determining the success of the project. Data was also to be collected in control neighborhoods.

Results:

The output from this effort would be of two kinds: short-term reports on project execution, and a longer term evaluation concerned with the achievement of project objectives. It was expected that by June, 1976 a first report would be available on characteristics of the project population and the initial conditions under which it would be settled in the project area. This report would be based on data from the baseline census of all applicants, already applied to the first large group [5176 households] which would also be used to defined the sample for subsequent work. Six-monthly reports on research findings of the evaluation, as well as operational reports designed to answer

^{1/} Such background information was much less extensive, however, than in El Salvador. In terms of the availability of such background, the four countries ranked as follows: the Philippines, El Salvador, Zambia, Senegal.

requests made by the executing agency, would be submitted to the Bank and IDRC.

Status:

Although the general research methodology would be based on periodic re-sampling, it was agreed that a stable sample of families should be followed longitudinally to note the progress of housing consolidation. Social and economic data would be collected on these families over time, to learn the reasons for variations in the rate of house consolidation. The baseline questionnaire applied to all project applicants would allow for a high degree of stratification of the samples drawn for the first application of different surveys, and future applications could, as a result, be simplified. Then, instead of trying to follow the same families and handling problems of attrition, it would be possible to draw new samples that would match the stratification of those originally chosen. Even after this approach was agreed upon, considerably more work would be necessary on the specific content of research instruments.

The proposed schedule of activities for the first year was accepted. Reports would be put forward on the socio-economic profiles of accepted candidates for project participation, their housing conditions prior to relocation, and progress of construction on the site.

III. Zambia

Dimensions:

The design of the Zambia project was substantially different from those in El Salvador and Senegal. Whilst part of the program included sites and services, its emphasis was on squatter area upgrading. Almost one half the low-income population of Lusaka, living in five major squatter settlements, would be involved, with each community receiving a number of communal standpipes, improved roads, schools and health clinics. Individual families would receive leases for their houses and surrounding grounds, and small loans and technical assistance to help improve their homes. Those families in the path of new roads would be given new lots in "overspill" means immediately adjacent to the squatter areas, and so remain members of their old communities.

The emphasis on area upgrading shifted the orientation of the evaluation from exclusive consideration of the creation of new environments to assessment of the ways existing neighborhoods change with improvements in infrastructure and services. It therefore became important to understand what local groups contribute to a community's adjustment to a project, as well as how communities make decisions on the placement of infrastructure improvements.

Here too it would be important to follow closely changes in residents' incomes and employment, but for somewhat different reasons than in the cases of the other two projects. With no significant amount of relocation taking place, there was no reason to expect the short-term

strains on a family's employment patterns that might pose problems in the other two projects, particularly in Senegal. However, there could be no minimum income standards for project participants in Zambia. All residents of a community chosen for upgrading would be involved in the project and would therefore be expected to pay for their share of improvements. It would be crucially important to establish the feasibility of this approach.

Data and Evaluation Activities:

The Zambian design was at the outset less fully articulated than the previous two had been. This was partly a result of the later recruitment of the primary researcher and other evaluation staff, but mainly the result of the timetable of project execution which (for an upgrading project with population in place) started with less lag. These circumstances forced the evaluation team to begin some field operations immediately, to avoid losing critical data.

The primary sample (or baseline survey) was to be the principal research instrument. Applied to several hundred families in each of the main communities to be upgraded, it would include basic social, economic, and housing history data. Additional questionnaires, applied to subsamples of the primary group, would deal with the magnitudes and rates of change in housing consolidation, health care, and household budgets. Extensive use of case studies was planned to help assess social changes over time.

The Zambia evaluation has evolved along lines similar to the others' with few exceptional aspects. It enjoyed the early advantage of

a very comprehensive prior study of the city, including all its squatter areas, by a group of British social scientists. ^{1/} Later, the broad coverage of the project was found to increase greatly the difficulty of finding appropriate control groups.

Results:

In the area of outputs, too, fewer advances had been made in preparation stages here than in the other project situations. In addition to quarterly progress reports, it was decided that the main results of the evaluation should be published when ready as working papers, and made available to the Housing Project Unit, the Ministry of Local Government and Housing, the Bank and the IDRC.

Status:

Over the following several months, the evaluation team developed the research design further, as additional survey instruments were prepared. As the team was completing the baseline survey for the community of George, it was also preparing to re-survey Chawama, where upgrading had been going on for the past year. ^{2/} The primary researcher noted the difficulty of finding qualified interviewers, and the fact that more staff time had been spent in direct supervision of interviews than had been anticipated. This was an early indication that

^{1/} National Housing Authority, "Lusaka Sites and Services Project," Lusaka, July 1973; Development Planning Unit Study, University College, London, 1974. See also David Pasteur, The Management of Squatter Upgrading: A Case Study of Organization, Procedure, and Participation, Saxon House, England, 1979.

^{2/} Original baseline data for Chawama would be taken from a 1974 study by the Development Planning Unit, University College, London.

staff resources might be insufficient to process and analyze data, and that the provisional schedule of activities might be affected.

IV. The Philippines

Dimensions:

The Philippines project embodied a number of difficult and important political decisions by the Government. With its commitment to improving the Tondo Foreshore area, the Government formed a National Housing Authority, replacing a network of ineffective agencies, to implement the project; adopted upgrading of squatter settlements as an instrument of policy; and gave new priority to shelter issues. This was no pilot project. It involved some 240 hectares of land in central Manila and about 200,000 people. Failure would be disastrous. Success could impart momentum to efforts throughout the world to promote effective policies to improve living conditions for the world's urban poor. For these reasons and others, both the NHA and the Bank were committed to a substantial program of evaluation from the outset. ^{1/}

Data and Evaluation Activities

In this case, the compromise between operation and research needs and personnel was much more carefully worked out than in the others. A clear distinction was created between process and impact evaluation objectives and activities.

^{1/} The most important additional reason was that this first "urban" project was developed deliberately out of as comprehensive and careful a consideration of Manila's overall needs as could be managed at that time. This level of address to the situation was pursued at the strong urging of the Bank

Process evaluation would seek answers to such pressing operational questions as the appropriateness of lot, layout, core unit, and house designs; the extent and impacts of reblocking during infrastructure installation in upgrading areas; the impacts of various forms of relocation and of community involvement in them; the adequacy of preparations for household movements; attitudes towards individual water and sewer connections, under individual and group metering; effectiveness of bulk water metering and shared sewerage facilities, where these are provided; the use of self-help versus contracted labor in house completion; the appropriateness of levels of standards for roads, footpaths, water, sewerage, community facilities, social programs, etc.

Initially, these questions would be pursued using financial and physical information generated during project implementation, as well as observation and "unstructured" interview techniques, including an attempt to collect information on "attitudes." As the research progressed, these processes would become more useful for their capacity to generate hypotheses, and for more rigorous testing within the framework of impact evaluation.

Impact evaluation would be oriented to producing information on the following kinds of questions and their interrelationships: changes in family income and in expenditures on housing, transport, and other goods; changes in labor force participation, employment levels and patterns; occupancy turnover in units, and income classes of occupants moving out and in; reaction to land tenure arrangements for accomodating

lowest-income families; changes in health and nutrition in the project area; demand for rental units; success in promotion of small industries and contracting businesses, etc.

These issues would be probed using statistical techniques to the extent possible. The main survey instrument, to measure large and general changes, would be a broad socio-economic questionnaire. This would be applied periodically (at least three times, at intervals of at least one year) to a sample of the project population and to suitably chosen control groups. Specific surveys, pursuing a broader range of questions on more specific topics (such as house consolidation or health) would later be applied to the same samples or to subsamples drawn from them. The decision to conduct a detailed survey of household income and expenditures was an important innovation. This survey would make use of daily and weekly diaries over an extended period, for a sample of 100 households.

The prescribed analyses would be supported by an exceptionally rich fund of socio-economic data and information on the subject population, the control groups, and the Metropolitan Manila population generally. Basic census and statistical information was already good and improving. Numerous surveys, and some re-surveys, of low-income populations and neighborhoods were available. These surveys provided

many valuable insights, and ranged from the impressionistic to the rigorous. ^{1/}

Administratively, it was decided at the outset that the evaluation unit, called the Research and Analysis Division (RAD) should be created as a permanent component of the NHA. Because of the number and quality of trained social scientists in the Philippines, RAD was created and soon staffed entirely with Filipinos.

Results:

By the time RAD was constituted, the Bank was no longer requiring periodic reports from the other evaluation teams. After a couple of quarters, it became clear that this was not an efficient means of reporting substantial information, and the requirement was curtailed to include only financial and administrative information. The requirement was dropped altogether once it was seen that the Bank's normal supervision activities were fulfilling these functions. Thus it was agreed that reports on the Philippines project, whether of operations or research, should be self-contained and disseminated whenever ready. Some consideration was given to format and to clearance processes. The former was a subject of some uncertainty, whereas the latter were readily agreed upon. Several procedures, including

^{1/} Among the latter was a detailed socio-economic survey of a 10% sample of the Tondo Foreshore population designed by the Urban Research Division and conducted as part of project preparations in 1974. But when it came time to use the information, it was discovered that the data had been lost, due to faulty recording of the sole tape or to its improper storage. Only a report making very partial use of the data survived. This unfortunate mishap points up the importance of appropriate and well administered data collection, cleaning, storage and documentation activities.

executive summaries for the NHA management, were tested in monitoring results.

The Evaluation Prospect at the Outset

A technical conference on evaluation of the three initial projects was held in Washington, D.C., January 12th through 17th, 1976. The conference, jointly sponsored by the Bank and the IDRC, was attended by representatives from El Salvador, Senegal, and Zambia. Also present were the primary expatriate researchers working in each country, representatives of the IDRC, and outside experts on certain subjects, such as health. As the meetings were held at Bank headquarters, large numbers of staff from the Projects Department and the Resource Division attended sessions.

This section reviews the conceptual and procedural matters agreed upon at that time, as well as considerations left unresolved. The commentary serves to summarize our expectations of the evaluation efforts at the time field work was getting underway. The material is presented under seven headings: housing consolidation, health, infrastructure, employment and income generation, finance and credit, community change and development, and general methodology.

1. Housing Consolidation

The projects were designed to promote rapid provision of better housing for those unable to find it in the existing market, to accommodate families on serviced plots as quickly as possible, and to minimize housing construction costs. These objectives, and the need to evaluate their achievement, were outlined in a detailed brief, and

accepted at the conference with one clarification. The long-held assumption that minimizing housing construction costs would demand a maximum of self-help and mutual-help labor was regarded as a hypothesis to be tested. If few household members were underemployed, and the opportunity costs of heavy participation in self-help labor were high, it might not be in a family's interest to maximize its own labor contribution. By defining the objective of minimizing construction costs broadly enough to include opportunity costs, project designs could reflect the need to optimize--not necessarily maximize--self-help.

In an effort to permit comparison among the projects, five specific measures of housing consolidation were considered: 1) the area covered by a roof; 2) the area enclosed by four walls; 3) the number of rooms, with an approximation of size and a note concerning their use; 4) the quality of materials; ^{1/} and, (5) estimates of the value of materials and labor used.

The researchers generally agreed that the evaluation teams would attempt to collect these data, though the approaches and methods of collection might vary. In El Salvador, for example, it might not be necessary to measure square meters of enclosed space directly, as room size and shape were largely determined by plot size and the core unit's position. Thus, the area enclosed might be inferred from the number of rooms.

^{1/} The UN had established three rough grades of material: temporary, semi-permanent, and permanent, but some further detail, such as distinctions between sun-dried brick, soil-cement blocks, and cement blocks, would add importantly to the analysis.

It was deemed possible to measure the value of houses built in the projects in three different ways. The first was accounting cost: the actual cash outlays required for mortgage payments, materials, and contract labor. The second was accounting cost plus an imputed value for self-help labor. The third was direct estimation of the market value of the housing produced. This calculation would comprise the accounting cost, plus the imputed values of any inputs "in kind," plus a profit element.

The first two types of valuation should not be difficult to calculate. Records were available from materials stores, families could be asked about amounts and values of sub-contracted labor used, and an imputed value of self-help labor could be figured in several possible ways, such as the generally acceptable shadow wage, if there was one, or the wage rate paid to construction laborers for similar tasks. It would also be necessary to estimate the time family members spent donating mutual- or self-help labor. In some cases, information on the second value might be available from other sources. ^{1/} In Zambia, for instance, a strict government land policy required that the sales price of a house and lot reflect only the improvements actually made on the lot and the house itself. In El Salvador, the executing agency set a policy that a family leaving the project within the first five years must resell its home to the Fundacion at a price equal to its actual

^{1/} Accurate evaluation of a family's opportunity cost for participating in self-help could come only after detailed consideration of family members' overall levels of employment and income, and assessment of the real opportunities that had to be foregone.

investment. At the same time, these considerations would increase the difficulty of estimating the third (market) measure of housing value by seriously restricting the recording of actual sales prices.

Two additional issues were discussed at the technical conference, but not resolved. First, it was thought desirable to establish a consolidated housing index that would bring together the suggested measures and permit a quick overall assessment of construction progress. Second, it was suggested that each project establish benchmarks against which the progress of housing consolidation could be measured. For example, on the basis of the appraisal report and the expectations of the executing agency, it should be possible to set a time interval for the completion of a limited core house. Where expectations of housing progress were not so clearly defined, as in most upgrading situations, it should be possible at a given time to identify those households not in line with the mean rate of progress.

It was expected that progress indicators would enable the evaluations to develop explanations for variances in housing consolidation rates. ^{1/} Such explanations were expected to fall into two categories: those related to project components, suggesting possible design modifications; and those related to exogenous factors, suggesting more fundamental difficulties or advantages of policy as a whole. The first category would embrace such factors as mortgage and

^{1/} A slow rate of development would not in itself imply that a particular family had failed in the program: the household might simply be husbanding its resources carefully. Similarly, a family expanding its home rapidly might be overextending itself financially in the process.

credit arrangements, types of materials used and their accessibility, initial level of core construction, technical assistance, tenancy regulations, physical design of the site, and location of the plot within the site. Community involvement or cohesion might also be a factor. Though not a component of the projects as such, community attitudes toward each project would be determined in a degree by overall design and the extent to which residents believed their experience had been considered in the design process.

The second category of variables affecting housing consolidation included pre-existing social and economic conditions of families, cyclical employment, availability of self-help labor, the magnitude of excess demand for housing in the city, and the overall physical conditions of the site.

The effort to understand the impacts of exogenous factors would require at least the collection of income and employment data over time for those households involved in the housing consolidation study. Such data could be collected by means of a panel survey, carried out every few months, and falling between applications of the general socio-economic questionnaire. Using a short questionnaire, the pane survey would register patterns of change in housing consolidation, as related to survey in household composition, income, and employment. The more rapid the rate of building, the more frequently the data should be collected, possibly as often as every two or three months, to insure accuracy.

The ability of researchers to analyze the impacts of the components related to housing consolidation would vary from project to project, depending on two major factors. The first would be the extent to which different levels of a given component were represented in the project, the greater validity of one level being used as a control for the other. The El Salvador project, for example, had clearly differing levels of initial core construction, whereas the Senegal project had none and the Zambia project was almost entirely devoted to upgrading. Absence of such differences in a project did not preclude the possibility of analyzing a component; however, their presence would clearly permit more rigorous analysis. The second factor was the extent to which accurate information could be collected about specific components: technical assistance inputs and the level of community cohesion would, for example, remain inherently difficult to evaluate.

2. Health

The issue that received most attention at the technical conference was evaluation of changes in health. Health-related goals of the projects had been stated only recently and in quite general terms. As a large scale health survey could easily absorb the entire resources available for evaluation, it was imperative to set priorities and reach clear understandings on what was feasible. Ideally, a complete health survey would operate at three levels: 1) identifying major health-related components and evaluating their functioning; 2) following the more important changes in health status; and 3) attempting to arrive at causal connections with components. It was generally agreed that the

third level of analysis would be beyond the scope of the present evaluation effort, although it might be an added research element at a later stage. The possible reasons for overall changes in health status would simply be too many and too elaborately interconnected for the limited number of controls available. ^{1/}

It was still possible to consider other levels at which a sound health surveys might function. Evaluation of inputs to the health system, and of the processes of their use, was regarded as potentially very useful to operational staff. It would call for measuring the effectiveness of the components representing major improvements in public health conditions. Rather crude analysis might contribute importantly to improvements in design: is the water pure at the standpipes? If so, is it still pure when it is used in the home? Is the sanitation system used and maintained properly? Is surface water drained quickly?

It was thought, somewhat mistakenly, that information for this level of evaluation would generally be fairly easy and inexpensive to collect. For example, the local health ministry could cooperate in

^{1/} It is helpful to cite an example. It may be found, for example, that a project's water reticulation system does provide dwellings with fresh water, and that the sanitation system functions well. However, it would be difficult to establish that one or more components (or service levels) had been most effective in, say, decreasing the incidence of child mortality. Some analysis may be possible if a project site has more than one level of service for a given component or series of components. Where this is the case, though, such as in Zambia (where there are pit latrines for upgraded areas and water-borne sewerage in sites and services areas), special attention must be paid to the selection of control groups. Those living in the sites and services areas are likely to have higher incomes and better health at the outset.

testing water purity in homes, and much of the data on use and functioning of other infrastructure components could come from large case studies and the researchers' informal observations. These data would also indicate residents' level of awareness of health concerns, and thus provide information useful for future project design on the extent to which a health education program could heighten a community's understanding of health issues, and on the direction such a program should take.

Two of the three initial projects (Zambia and Senegal) included construction of on-site clinics. To help in the planning of these facilities, it would be important to collect baseline data on residents' current experiences with health services (i.e., from whom and how often services are sought). Within a few years after the establishment of the clinics, a specific survey might be warranted. Such a survey would seek to determine whether the facility provides the services the community requires (outputs) and if residents are taking full advantage of the services offered (effects).

Researchers agreed that an attempt should be made to measure major changes in health status, understanding that the severe constraints on such analysis might mean poor results. Noticeable changes might occur, but participants were warned not to expect dramatic shifts in health status within the first few years. ^{1/} The medical staff also stressed that knowledge about health inputs would be prerequisite to establishing a context for analyzing changes.

^{1/} In this area of research, a very carefully drawn control group would be needed to permit distinction of the project's impacts from those of broader exogenous events, such as the general improvement in the city's income levels or environmental quality.

With these points understood, the following guidelines were adopted for establishing measures of important changes in health status: (1) The program should be adapted to the specific country situation, there being no standardized series of indicators appropriate to all circumstances. (2) The study should focus on infants and/or young children: these groups are not only likely to respond most quickly to improved health conditions, they are also the ones most confined to the environment of the project site. Adults may well spend most of their days removed from the community in places of employment. (3) House to house surveys must be conducted, as data from clinics on frequency and reasons for visits cannot be taken as an accurate sampling of the health status of the community. (4) The research should remain on as simple a plan as possible, with no more than 3 to 5 measures adopted as an adequate surrogate for overall health status profiles. Highly trained personnel and special equipment should not be required, and the tests and questions should be as simple as possible to minimize ambiguities.

Within these guidelines, some major indicators were proposed: (1) Number of deaths, among children under two, due to diarrhea and respiratory diseases. It was assumed that most parents would have some knowledge of the cause of a child's death, if only the symptoms suffered. Diarrhea and respiratory ailments are the two most frequent causes of infant mortality in squatter communities, with death rates of 50 to 60 per thousand live births from each disease. (2) A second indicator would be the incidence of diarrhea and respiratory

diseases. It was assumed that most parents would have some knowledge of the cause of a child's death, if only the symptoms suffered. Diarrhea and respiratory ailments are the two most frequent causes of infant mortality in squatter communities, with death rates of 50 to 60 per thousand live births from each disease. (2) A second indicator would be the incidence of diarrhea and respiratory diseases among children under four, measured as the number of new cases over a period of time. As with the mortality data, this information would be garnered from household interviews. (3) Anthropometric measurement: weight for height. It was deemed crucial to keep records on those small children who showed no signs of severe malnutrition. Although there was no nutrition component in these projects, a lower incidence of debilitating disease, better housing and higher incomes should over time have impacts on children's levels of nutrition. (4) The incidence of tuberculosis, to be determined by a simple skin test; and (5) the incidence of certain skin diseases easily recognizable by paramedics.

Mortality figures are generally the most reliable basic health data: the fact of death is beyond dispute, and a parent's recall of a child's death is likely to be accurate. However, whilst the question is easily asked, the sample size needed to gather such data tends to be very large relative to the scope of the evaluation. It might be necessary to visit well over one thousand families in order to collect a sufficient amount of data. However, in Dakar, there would be fewer than one thousand families living on the site for a long time. In El

norm: what is regarded as sickness in an economically developed country may be seen as normal in a poorer one.

Other possible questions bearing on residents' perceptions of ill health further illustrated the difficulty of research in this area. Whether a person's reason for missing a day of work is or is not due to sickness may depend as much on his or her felt need for job security and for the day's wages as on actual feelings of illness.

Additional measures of health status might include collection and analysis of blood, stool, and urine samples. Sample sizes in such cases could be relatively small and still yield accurate results with adequate analysis. The proper level of analysis might be effectively carried out if the evaluation team could secure the assistance of a health agency in the field and gain access to a laboratory. ^{1/} In general, the extent of this level of health research should depend on the available advice of qualified medical personnel in each country, and on a careful comparison of the costs with the expected reliability of the data to be collected.

3. Infrastructure:

Ideally, it is desirable to know how and to what extent infrastructure contributes to the realization of project objectives, and what levels of services are minimum prerequisites, or most cost effective, for attaining these results. Again, establishing causal

^{1/} It is clearly the obligation of the research team to report the causes of severe illness to the health authorities. Provided the affected individuals are then helped, it would be necessary for follow-up studies to be carried out on a freshly drawn random sample.

3. Infrastructure:

Ideally, it is desirable to know how and to what extent infrastructure contributes to the realization of project objectives, and what levels of services are minimum prerequisites, or most cost effective, for attaining these results. Again, establishing causal links between infrastructure components and project impacts would be very difficult. However, data on the efficiency of infrastructure should be relatively easy to obtain, and are directly linked to design decisions.

It was decided that inquiry should be focused on four aspects of infrastructure provision and use. The first is an historical record of infrastructure provision: When are services provided and when do residents connect them? Such data should be readily available from project records, or easily obtainable by means of a questionnaire. The second key aspect is affordability; here too most important information should be available from the executing agency. The third aspect is acceptability; case studies and informal observations should reveal community response to infrastructure maintenance and use. Finally, households' ranking of the various infrastructure services would be an important area of inquiry, although it was not deemed a feasible proximate objective of these evaluations. What should be feasible would be to discover the indicated directions of change, whether specific service levels should be increased or decreased, eliminated or newly instituted. Most likely, these data should be obtained principally from

poverty, even a minor disruption of earnings could cause financial strain and possible default on mortgages and other obligations. Even though employment generation was not a specific objective of these early projects, the employment impacts of project-related activities had to be considered.

The general socio-economic questionnaire would be the principal instrument for collecting information on changes in residents' income and employment patterns, and on the effects of relocation, where relevant. Left unresolved at the technical conference was the issue of the depth of analysis that should be attempted. The most accurate and complete data would come from interviews with individual household members, ascertaining amounts and sources of each member's income. A time budget for the working and commuting hours of the day would also be necessary for careful measurement of informal and secondary employment. The expense of collecting employment data in such detail would probably be prohibitive. And so another set of trade-offs would develop between reduced sample size and simplifications in the data-gathering process, such as interviewing only the major household income earners and asking them about the rest of the family's earnings.

Since employment generation was to be a component of future housing project designs, several areas of research were considered to develop an understanding of the employment generation process. The first measurement required would be the easiest to obtain: the total number of jobs generated directly from site preparation and construction of infrastructure. Data should be available from the contractors and

could simply be aggregated. It would also be useful to know the secondary and tertiary impacts of construction, such as the number of jobs created in sub-contracting and in the various industries supplying goods to the site. But whilst some measures of such indirect effects might be readily available, an effort to track them down was considered beyond the scope of the evaluation.

The approach taken to generation and support of small business activities varied from project to project. In Zambia, it was expected that the existing small businesses within the squatter areas would stay and service the communities. However, it would be important to measure any change in their numbers and trade volumes due to increases in income of the project population. In Senegal, lots had been designated for allocating or developing them. In El Salvador, a small fund and some technical assistance were available to aid the establishment and growth of community business cooperatives.

In all countries, data could be most readily collected by compiling an inventory, updated occasionally, of the businesses in and at the periphery of the community, noting their activities, the number of people employed (by skill level and/or wages, if possible), and the total investment in the enterprise. The reasons for change in the

inventory--new businesses opening, others closing--would also be recorded. ^{1/}

Although all three projects have the use of self-help labor as a premise, it was recognized in advance that many families would opt for sub-contracting some or much of the work they might do themselves. The amount, kind, and value of this sub-contracted labor would be available from the questionnaire on housing consolidation.

5. Finance and Credit

The conference's consideration of finance and credit centered on four major issues: residents' access to credit, their use of it, their ability to afford it, and their realization of the full value of the new assets. The researchers believed that many of the questions in this area posed difficult data collection problems. However, a plan of inquiry was formed.

The first major objective of the credit components of the projects was to improve access to credit for low-income families. It would be important to know whether there are other forms of credit available to these families at rates they can afford. The researchers felt they would have no difficulty describing the processes of credit utilization: how many families require credit for what stated

^{1/} There is a methodological problem implicit here. The sites and services projects could be appealing generally to the most able and upwardly mobile of the poor, and thus draw the most capable entrepreneurs as well. If there was a rapid development of small businesses, it would not be clear whether this was a result of the kind of applicants interested in and selected for the program. This is an example of the broad methodological difficulty of distinguishing between the effects of the project on the population, and those of the population on the project's success.

purposes. In all three initial projects, records from the executing agency should be able to provide most of the relevant data. However, determining what other sources of credit are available, at what levels, under what circumstances, may be a difficult matter, particularly in Zambia and Senegal. The researchers averred that the alternative financial channels would require extensive studies in their own right. They suggested further that residents would be unwilling to disclose their total financial resources and transactions, and that the interviewers, by asking such questions would appear to be credit inspectors, and so jeopardize community goodwill toward the evaluation.

While not discounting these problems, Bank operational staff believed that whatever information could be collected on project residents' alternative sources of credit would be of considerable value. How much information of this kind could be collected would depend on the cultural context, and on how questions were posed. Researchers suggested that credit questions could be asked most amenably in the context of the housing consolidation survey, where the availability of resources would naturally be an issue.

The early credit schemes tightly limited funds and carefully proscribed their uses, targeting them primarily for mortgages, materials loans and loans for small businesses. Research must discover whether these regulations and procedures proved efficient in furthering housing consolidation and business expansion. Perhaps the most difficult question to address would be whether, and to what extent, the credit lines (amounts and conditions) set for the projects impeded

housing consolidation. The credit limits and regulations presupposed the possibility of abuses by project beneficiaries. However, this was not deemed a promising area for research, given residents' natural unwillingness to discuss errant behavior with outsiders.

What could be useful were residents' reactions to the credit programs. Do they feel sufficiently secure with their form of tenure to invest? Is there a demand for alternative mortgage repayment schedules? Do people unwilling or unable to perform their own construction need additional credit to hire contractors? How well do the systems for services fee collection work? Useful information these questions often come, in the first instance, from the case studies and other less structured research instruments.

A crucial feature of credit programs was their role, making the housing scheme affordable to target populations, a key condition for project replicability. With access to the executing agencies, records on payment and default rates on loans and services charges, plus survey data on income and family employment, the evaluation teams would be able to delve deeply into the issue of affordability.

Finally, it would be important to know how project residents responded to their new assets, both credit and land tenure. Because they would be receiving urban sites with tenure at a price far below full market value, it would be interesting to know how beneficiaries treated this windfall. Did they use it as collateral for additional loan? Did they rent part or all of their space? Or did they sell the property and realize a capital gain?

6. Community Change and Development

It was widely accepted that the successful completion of construction in self-help housing projects depended upon residents' positive attitudes toward the project and a high level of participation. The corollary assumption was that construction processes and related community activities would strengthen local organizations and increase individual's involvement in community matters. This process of "community development," as it has been called, remained difficult to define or to measure. The technical conference did make some headway in clarifying directions for research.

First, it was recognized that the term, "community development," might be seen as tendentious. It does not distinguish the experience of residents in sites and services projects from that of squatter upgrading project dwellers. For sites and services, it would be strictly accurate to speak of community development as community formation: residents formerly scattered throughout the city are set the task of forming a new social entity in their newly shared setting. If applied to a squatter area upgrading project, however, the term "community development" might imply that the existing community organizations are in some sense deficient and in need of improvement. A more accurate phrase might be "community action," referring to the residents' collective need to adjust to the new presence of the project by adapting existing social organizations or forming new ones.

Defining what is measurable in this sphere is no easier than defining the issue itself. ^{1/} The most valuable research would be a clear chronicling of the community's experience as it grows and adjusts to the project. The first report in such a chronicle would serve as a baseline study, indicating the structure of the community at the beginning of the project, to which future reports might be compared. The chronicle would have to include such questions as: what organizations are active within the community, and which appear most helpful in the housing consolidation process? How do residents evaluate the assistance rendered by social workers and community development teams? The level of community participation in organizations, and the extent of a group's authority should also be considered. It was expected that the community, particularly in upgrading areas, would modify the project cooperatively with operations staff, rather than simply accept its initial definition of individual and collective needs. The evaluation would attempt to ascertain the nature and intensity of community involvement in project planning.

Most of the information pertinent to these questions never come from case studies, interviewers' observations, and interviews with community leaders. A moderate emphasis on evaluation of community change appeared appropriate to participants from Zambia and Senegal, and

^{1/} One can easily ask attitudinal questions in a survey, but usually not to great effect. A few minutes' interchange over a questionnaire does not permit development of the rapport between interviewer and respondent necessary for information to be understood and accurately noted. However, a survey can be used to collect such simple data as the organizations to which a resident belongs, and with what degree of involvement.

to Bank staff. However, the Fundacion in El Salvador gave much greater emphasis to questions of social attitudes and organizations, involving the evaluation team in extensive research on these issues. It was not expected that the added attention to these issues would have a high opportunity cost for other areas of the evaluation, because the Fundacion researchers expected to have the assistance of affiliated social workers in administering many of the less formal research instruments.

7. General Methodology

This section reviews methodological concerns common to most of the research issues covered during the technical conference.

Accuracy:

Bank staff agreed there should be no pre-determined standards of accuracy for the evaluations. The degree of accuracy would necessarily vary with the ease of data collection and the kinds of instruments used. For the major surveys, collecting objective data on such topics as employment, income and housing consolidation, standards should be fairly high, and sample sizes chosen accordingly. In the less structured research procedures, information gathered could only be indicative of patterns of change. No fixed degree of accuracy can be established.

Controls:

Evaluating the fulfillment of project objectives would require a quasi-experimental design and the use of carefully drawn control groups. The ideal sample frame for such control groups would be

clusters of households from communities chosen for their comparability with project populations in such terms as income, family size, and past housing experience. This procedure would make possible comparison of individuals within the project and those outside it, and comparison of the project as a community with other communities. This community level comparison would be important in evaluating issues such as the functioning of organizations.

Finding such perfect control groups would be prohibitively expensive and difficult, if it could be accomplished at all. Because city-wide census data was insufficient in all three project locations, it was in fact very difficult to determine which communities were usefully similar to the community of each project. Following non-residents over time would pose even greater problems: the rate of turnover was likely to be high, and respondents would have no particular incentive to cooperate with the research process or to give accurate answers.

It might be enough to find a control group that paralleled only the socio-economic characteristics of the project population. One would then note the degrees of differences between the two communities for the appropriate variables, and see how these changed over time. Another possibility considered at the conference was to look for discrete controls to be used for specific topics, rather than for the entire range of issues under evaluation.

The techniques for coping with attrition from a sample all require additional expenses which mount quickly unless one assumes,

perhaps incorrectly, a relatively high level of community stability. One way to adjust to constraints on time and resources is to draw a new random sample of households from the control communities each time a survey is repeated. Limiting the controls to following communities, not individuals, would preclude detailed study of the changing circumstances of particular families in each group. However, it would still be possible to compare aggregate changes in incomes, employment, and other variables.

Even with these ways of dealing with comparability and attrition, establishing a reliable control was still likely to be difficult. The difficulty would be extreme in the case of Zambia, where some 80 percent of Lusaka's squatter population would be involved in the project. The other 20 percent was scattered in a series of small communities demonstrably different from the experimental communities. In this case, it might be appropriate to develop controls internal to the project, depending on the range of differences between the communities involved and the differences in the mix of components.

Case Studies:

Case studies serve three primary purposes: answering questions that more structured research instruments cannot handle, enriching "harder" quantitative data, and generating hypotheses for further testing. In this book, the term "case studies" is used to cover all "unstructured" research techniques, including panel surveys focused on a limited range of issues, casual observation by interviewers, and more formal observation guides.

Three principal types of case studies were discussed at the technical conference. The first represents the common understanding of the term, in depth studies of typical individual households. The second type focuses on families that are of interest because of the ways their experience deviates from the norm, demonstrating the limits of change caused by the project. The third could more properly be called the use of informants. Here one would search out particularly well-informed and articulate members of the community and use their observations in compiling an oral history of community changes.

For case study data to be useful, the respondents must be reasonably well-spoken and willing to participate, a clear potential source of bias. Researchers must therefore know the background interests and attitudes of respondents, and appear as neutral visitors to the community, to be able to solicit comments from a large number of individuals representing various viewpoints.

Such informal data collection techniques as written chronicles based on interviewers' observations would be very useful in El Salvador, where social workers involved in the projects could easily record their observations in systematic fashion. There is some risk of bias in having members of the executing agency participate so directly in the evaluation. However, this was not thought to be a problem in El Salvador, as staff there appeared so anxious to achieve a careful and accurate evaluation of project activities.

Standardized Measurements:

Comparison of research results from the three projects was an original and long-standing goal of the evaluation effort. However, it was soon realized that such comparisons could not be easily made, for the projects differed considerably in both design and cultural contexts. Accordingly it was decided that priority should be given to designing the most appropriate evaluation for each set of circumstances, even if this meant compromising the possibility of comparisons. The assumption was that the most useful comparisons would anyway be those describing general directions, rates and magnitudes of change, not those at the level of specific indicators.

With the understanding that data bases should be kept as comparable as field conditions would permit, four possible areas of measurement standardization were considered:

1) Income: It was agreed that data should be recorded directly onto questionnaires and subsequently onto cards in raw form. Preceding into categories would severely limit researchers' ability to manipulate the data later. When processed, the information would of course have to be broken into categories according to the analysis contemplated. Most available computer packages would both retain raw data and create a new variable for the pre-coded, categorized data.

2) Health: The relevant health indicators have already been discussed. Before determining which of them would finally be used, researchers would have to consult with local medical personnel on the feasibility and reliability of each measurement.

3) Occupations: As with income, information on occupations was to be initially recorded without precoding. Interviewers would have to be given careful instructions about how to describe and record each respondent's type of employment. In this case, it would be necessary to code occupations into categories, at however low a level of aggregation, to prepare the data for computer processing. As the Bank had no preconceived hypotheses demanding the establishment of a particular set of occupational codes, it was decided that researchers should design the categories to suit their local contexts, subject to review by Bank staff.

4) Housing Consolidation: Here the Bank staff stressed arguments already presented for a reasonable standardization of measurements. The specific indicators to be used have been discussed above also.

Sample Attrition

Given the importance attached to the collection of longitudinal data for the experimental population, it was clear that resources would have to be committed to dealing with the problem of sample attrition, in both control and experimental groups. But, with the experimental groups, it would be necessary to avoid the kind of compromise resolution of the problem that might apply to control groups.

After much discussion, the following general and flexible method was adopted. An initial sample would be drawn and interviewed, somewhat larger than would be needed in the absence of attrition, to maintain confidence intervals over the initial evaluation period, even

if the expected attrition did occur. ^{1/} If some of the residents should leave the project, every effort would be made to note who they were and why they left. For future applications of the major surveys, three possible outcomes might be considered: 1) If the attrition rate is minimal, and those left are not from any clearly defined group, it might not be necessary to adjust the remaining sample at all. 2) If attrition is somewhat greater, families recently moved to the project site could be substituted for those who have left, making it necessary to carry out two levels of analysis. The first would be a longitudinal study of families that had remained in the project and on whom data had now been collected through two or more questionnaires. This sample would no longer be representative of the community, as the new residents might be quite different from those who left, but it would remain statistically valid in measuring change among residents who had stayed over time. The second analysis would not be longitudinal. It would include the data on the substituted sample of new residents, along with that on those remaining in the original sample, and would describe the social, economic, and housing situations for the community as a whole. Comparing these results to the original survey would reveal how the

^{1/} This method would be applied to the El Salvador and Zambia projects where relatively small populations in several locations would have to be surveyed. In the case of Senegal, the process would be simpler in that the entire population of 14,000 families would be in one location and would have been covered by the baseline survey administered in the selection process. Thus, longitudinal analysis of the project population as a whole could be based on periodic random samples. Where it was deemed necessary to "follow" a particular set of families (e.g., for health characteristics) or plots (for housing consolidation), this could be done through special surveys.

characteristics of the original community were shifting over time. 3)
If attrition is considerable over a few years, even the remaining sample
of original residents could become so small as to be statistically
questionable. A number of courses of action, too numerous to describe
here, would be possible. However, all had serious drawbacks in that
they would force a choice of which types of comparison would be most
important to make.

Chapter 8: DEVELOPMENT OF THE EVALUATION PROGRAM

Early Experience, 1976-77

Nearly all the expectations of evaluation progress raised by the Washington technical conference were soon dashed. The main reason was delays in project implementation, which were various and largely unforeseen. Before long, certain constraints on the evaluations created additional factors in the delays: the scarcity of properly trained staff, insufficient provisions for training, and severely limited data processing capabilities were among the more important factors. The earliest consequence of the resulting delays was to render the next conference, held in Dakar (October 25 through November 2, 1976), relatively unproductive. ^{1/}

The Dakar Conference

Though some expectations were frustrated, there were also accomplishments at Dakar. The El Salvador evaluation had progressed rapidly, despite slippage in certain sub-projects. Its results, including the identification of problem areas, provided a useful model for representatives from other countries where programs were progressing less rapidly because they were not as well defined and the evaluation teams not as strong. The conference also identified a potential problem with the El Salvador evaluation: it might become so broad as to lose sight of the principal operational goals of the program.

^{1/} This was the only conference (of seven held between July, 1975 and November 1980) that was not very productive. So, although there might have been some gains from postponing this particular conference by several months, scheduling the conferences approximately annually seems to have been a sound policy.

A conference at this early data was worthwhile because it led to communication and coordination between the research teams, and to the development of some specializations among them. In this instance, the joint learning aspect of the conference was enhanced by an associated field trip to projects in Zambia and a number of other African countries. This pleased the country participants, as did the initial acceptance of the invitation from the government of Senegal.

Perhaps the most important accomplishments of the conference were methodological. Crucially, economic analysis was reaffirmed as the central process in the evaluation effort. At the outset, the intention had clearly and explicitly been to rely primarily, though not exclusively, on the tools of economic analysis. This has remained an objective. However, those engaged to develop this approach and staff the field teams were not economists, but anthropologists, sociologists, geographers, and planners. At that time, few economists had developed either the motivation or the ability to collect reliable micro data in the field.

Between 1974 and 1976, many of us connected with the evaluation seemed to have lost sight of its methodological objectives. This may have begun during the early phase, when research was concerned solely with the Senegal project, the others not having been selected yet. This project encountered primarily non-economic problems during its early years: establishment of effective operations during year one; physical provision of infrastructure during year two; and initial selection of applicants during year three. The only questions with

economic content that were topical at that time pertained to applicants' incomes: To what extent did they fall within the target range? Should adjustments be made for family size? The majority of interesting economic questions attended the allocation to families of sites they could afford to develop, a process that had progressed little by the time of the Dakar conference.

It also seems likely that some slippage occurred during the first year of field work because of inadequate communication between supervisors and non-economists on the field teams. The slippage took the form of loss of specific detail. Hypotheses were formulated in economic terms, but the specific questions and observations in the field survey instruments did not reliably record information in forms or to scales that were conducive to economic analysis. This problem was discovered during scrutiny of the early reports, questionnaires, and study designs. Once the problem was identified, Bank staff explained it, with concrete examples, to evaluation personnel and steps were taken to resolve it.

Another significant methodological decision stemmed from an observation regarding sample size during the planning for the Philippines evaluation. The observation was that, in judging the statistical validity of a sample, accuracy is primarily a function of the absolute number of interviews. Thus, if the sample size necessary for statistical validity does not vary significantly with population

size, data gathering costs per household of population covered were minimized by choosing the largest continuous projects for analysis. ^{1/}

Such factors had not been considered in the planning of this pilot program, as other criteria for project selection had prevailed. Everyone's early thinking about research designs was conditioned by the experience of the Senegal project. Optimal research design and cost minimization did not appear to be in conflict in the Senegal project, which was designed to establish 14,000 families on a single continuous site. The importance of assessing research results against unit costs for data collection began to strike home only when the teams got down to planning details of the programs for El Salvador and Zambia, where the

^{1/} Here "continuous" refers both to the population covered and to the treatment. That is, the population should come from the same distribution, such that sub-population characteristics vary randomly; similarly, the treatment should be uniform, or vary randomly. Both matters may depend on circumstance, project design, or hypothesis. For example, there may be racially or economically determined enclaves within a geographically continuous area, or a project may be designed, for political or economic reasons, to cover several non-contiguous areas. Either factor is likely to fragment the project population into sub-groups requiring their own samples to the extent their differences are thought to matter. Similar considerations apply to different treatments (or packages of project components) and to control populations and conditions.

project populations were scattered in a number of locations covering different sized areas. ^{1/}

This finding was not of much immediate use, since the Senegal and Philippines projects had large, more or less continuous populations, and the target populations in Zambia and El Salvador had already been irrevocably chosen. Its use was in guiding resources at the margin to the larger sub-projects in the latter two countries. This principle would, however, prove valuable in the design of future evaluation systems.

There were other methodological clarifications made at the Dakar Conference. For example, there was rapid progress in

^{1/} The following is a simple model.

- Let: S = prescribed sample size (to give desired confidence intervals)
 p = proportion of population assumed to received treatment by time of first resurvey.
 a = assumed annual attrition rate
 n = number of years from first to last application of survey.
 x = cost per household interviewed of questionnaire preparation, interview and preparation of clean tape.
 R = number of application of survey.

$$\text{Total cost, } C = R \frac{xS}{p} (1 - a)^n$$

Unit cost, or cost per household in the total population (N):

$$c = \frac{C}{N} = R \cdot \frac{xS}{pN} (1 - a)^n$$

This formula can be used in two different ways. First, since none of the other variables varies systematically with population size, unit costs are lower the larger the population. This is of use only as a rough guide in allocating evaluation resources. Second, since most of the parameters (notably p, x, n and a) can be expected to vary -- even if not systematically -- with population size, the entire formula can assist in making some detailed design decisions.

developing means of measuring and recording the housing consolidation process. The discussions on health, on the other hand, may have been premature, as they did not yield the cautionary guidance it would have been helpful to have had at that time.

The results of the Dakar conference may not have justified its cost; but, then, there were no prior known examples of such activity on which to base expectations. The most unfortunate factor was perhaps the negative impression given operational personnel who saw the activity, surmised its costs, and had still to wait some months to see the kinds of results they had been expecting.

Dakar to Lusaka

Between the conference at Dakar and the one a year later at Lusaka, monitoring and evaluation activities within the Bank's urban sector were much reviewed. These reviews were sometimes formal, programmed assessments, sometimes ad hoc reactions to events. In large measure, they were responses to the growing impatience of operations personnel for study findings to aid in their decisions, and their disappointment with the slow pace of the evaluation program. 1/

The Research Division conducted the first annual review of the evaluation program after the Dakar Conference, as part of its normal management procedures. There was a joint staff review of monitoring programs for Projects Department management, and a joint management review of the evaluation program, both culminating in the third quarter

1/ This impatience did not discount for the slow pace of implementation of the project themselves.

of 1977, two years after the essential decision to proceed with such programs. These deliberations produced a number of interim findings and some important corrective actions.

The problems these reviews disclosed were closely inter-related: there had been inadequate provision for resources at the center; objectives and resources in the field were generally out of balance; there were management problems in the African programs; and production and dissemination of research results were already turning into formidable tasks. It came as no news to the Urban Research Division that there was a shortage of resources at the center to supervise field operations, provide advice and technical assistance, collaborate on analyses, and exercise quality control over output. The program had been promoted and launched, despite no prior commitment of staff, because the Division Chief and concerned staff believed in it. Repeated efforts to secure staff positions for this purpose had failed. Beyond the small amounts of professional staff time given to it during 1972 and 1973, this program had been managed by the Division Chief and a research assistant. These circumstances were still better than those of the monitoring programs managed by the Urban Projects Department. The brief review concluded in July 1977, advised departmental managements that to date the full potential of the associated monitoring had not yet been realized by the Bank. Despite the value of such work, the program was still in its infancy and suffered because the concerned department had not been willing or able to provide the resources needed to have the field data properly disseminated and

utilized. This assessment concluded with the recommendation that a full-time staff of at least one person be allocated to this end. 1/

There was an imbalance between program objectives and resources for their accomplishment in the field because, from the outset, research goals had been overly ambitious and resources budgeted for their achievement unrealistically low. Even in El Salvador, where field capacities were most favorable, the imbalance was felt. When the El Salvador pattern of research objectives and instruments was applied without fundamental modification to the circumstances of Senegal and Zambia, where constraints on this type of work were much more severe, even larger discrepancies appeared.

The transfer of management from expatriate primary researcher to Senegalese counterpart in June 1977, and planning for such a transition in Zambia in June, 1978, revealed deficiencies in the original arrangements, particularly with regard to these African countries. In retrospect, there had been inadequate planning for research direction, staff development and training, and eventual transition from expatriate to local management.

The process of transferring management responsibilities in these two countries happened also to reveal certain largely administrative shortcomings in past management by primary expatriate researchers. These were of three kinds: The first were administrative irregularities, such as the Research's Division approval of car allowances for field staff, contravening normal Bank procedures, on the

1/ This recommendation goes unheeded by the institution to this date.

grounds that no progress could be made otherwise. The second were administrative irregularities not sanctioned by the Research Division that were due to the ad hoc status of the evaluation teams ^{1/} and the primary researchers' lack of management experience. An example is salary determination. Finally, there were administrative deficiencies: poor management practices and poor reporting to the Bank and the IDRC of financial transactions. These problems were brought to light by Projects Department administrative officers, and were eventually resolved through the joint efforts of the Projects Department and the Research Division.

Limitations on what could be accomplished in the field, lack of resources at the center, and communications problems (due to distance) between them meant that analysis, report writing, and dissemination of evaluation findings proceeded more slowly than anticipated. The delays were also due in part to the variety of potentially interested audiences for research results and the Bank's failure to provide clear guidelines for choosing among them particularly in regard to dissemination.

The identification, evaluation and management discussion of these problems led to a set of decisions in the Fall of 1977 that bode well for the remainder of the trial period. More resources were provided at the center. The Urban Research Division was permitted

^{1/} That is, the fact that they were not fully integrated into the bureaucratic structures and thus governed by a specific set of regulations.

to hire a professional staff member to work full-time on the evaluations. This was a great step forward, even though the permission was granted somewhat hesitantly and implemented only after the Projects Department agreed to "provide safe haven" for the staff member after two or three years, in case the Development Economics Department eventually terminated its involvement in evaluation activities. In addition, the Research Division was permitted to continue and extend its recent practice of hiring contract personnel with its discretionary resources.

At the same time, the Bank and IDRC granted incremental additions to resources for field operations. Recommendations for transfer of field management to indigenous personnel, along with complementary provisions for technical assistance and support, were also approved, though most of the latter proved difficult to implement. Work programs for the ensuing year were approved, and revised criteria and procedures for reporting evaluation findings were accepted. ^{1/}

Lusaka Conference

Some time was spent tying up administrative details, but the preponderance of preparatory efforts for the Lusaka Conference was spent on the substance of the evaluations. This also held true during the Conference and afterwards.

As with previous and subsequent stock-taking exercises, evaluation results contributed to an improved understanding both of how

^{1/} However, this has remained an area of continual flux and experimentation. It is also interesting to note that this was the period in which the Philippines evaluation began, and involvement in Colombia and Indonesia also commenced.

the projects should be viewed and of how the evaluation work should be performed. The first category has to do with a more refined appreciation of project objectives and means. Interest in project components and incentives overlaps the second category: "findings" about the projects, which usually emerge as a result of iterative processes. Typically, there are initial indications of what is happening. These arise in any number of ways, ranging from informal, off-hand observations to the results of formal statistical tests. The latter were yielding more as time went on. The next step was to reformulate hypotheses and tests, which each subsequent round of findings resulted more and more from the formal statistical procedures. ^{1/}

Evaluation has often been criticized as being of little use because its findings are neither new or convincing. Both assertions miss the point; the second is fundamentally wrong.

What an evaluation should attempt to do is assist in the process of improving everyone's understanding of what projects can and cannot accomplish, and how efficiently they can be made to function. Thus evaluation contributes steadily to the processes of modifying current projects (as necessary) and improving the design of future ones. To expect evaluations to come up with something "new" is a misunderstanding. It overlooks the fact that the formulation of hypotheses for evaluation is typically a joint undertaking of

^{1/} The number of questions to be addressed is so great, the temptation is always to address new questions rather than refining knowledge about the old. This fact can work against the application of very many statistical tests.

operational and evaluation staff. The tests of the hypotheses may lead to their acceptance or rejection, or to ambiguous results. In no case will anything "new" be involved, in the sense that each possible outcome will have been previously stated, or at least implied.

Despite these facts, the evaluation program has come up with a fair number of unexpected findings. These have already been presented in Part II; however, a couple of early examples will serve to make the point. One thing that may happen is that relationships between variables emerge as important that had not been foreseen as important. An example would be the early findings of unintended bias against female-headed households in the Senegal projects' selection process. A second type of revelation may be that, whilst the direction of a result may not be surprising, its magnitude may be of a surprising order. An example would be the large proportion of contract labor (as opposed to self-help) used in the Zambia project.

The greatest usefulness of systematic evaluation, however, is in increasing one's confidence in findings, the more so as the research follows a rigorous model. This increased confidence is more important the further removed the person who must act on findings from first-hand familiarity with the project supplying the example; in short, rigor helps findings to stand the test of time, distance and changes of cultural and institutional setting. How convincing a particular evaluation is in practice will largely depend upon the establishment of priorities and the allocation of resources, for which operational staff are likely to be responsible.

A related set of questions concerns the derivation of findings: where do they come from and who should get the "credit"? Many considerations make this a relatively unproductive line of inquiry. Knowledge and understanding of project processes develop in so many different ways. Some changes may be generated by the evaluation itself. There is really no way to untangle all the various contributions to the research process, or to weight their relative importance. Though we try to explain how and why these processes have worked, in order to improve their workings in the future, we are convinced it would not be fruitful to assign credit for findings, even where this might be possible. We concentrate on the ways of deciding whether an evaluation has been done well; for only if it has been done well will it make contributions that justify the assignment of resources; or for that matter, give rise to bickering over who discovered what.

The first two years of evaluation efforts and review, culminating in the Lusaka conference, yielded three major kinds of outcomes: conceptual advances; initial findings; and reformulation of hypotheses, leading to modifications in designs.

Conceptual Advances:

With respect to conceptual advances, the greatest progress was made in two important areas: the first being the economic framework for analysis of project activity; the second being the more limited, but vital, areas of credit, tenure, subsidies, and affordability.

The Lusaka conference of project managers and researchers identified the salient economic issues as affordability, housing demand, and housing consolidation. These categories could be couched in the demand-supply framework of a market economy.

Policy makers, in this case project designers, can be characterized as viewing housing as a "merit good" that should be consumed at a level in excess of that determined by the market and at a price the low-income population can afford. The projects aim to produce this result by providing a package of benefits to stimulate consumption of housing services by the target income group. However, because the projects aim to satisfy "merit wants," they are designed to reconcile the expectations of policy makers and target populations with institutional constraints. Whether these expectations and constraints are in concert or in conflict is the subject of the research.

Figure 8.1 provides one way of interpreting urban development projects from the view point of affordability. ^{1/} The "typical" individual in a low income setting is supposed to be currently consuming a level of housing services which is below that considered desirable by policy makers. Making the usual assumptions about perfectly divisible markets, the situation of the "typical" household can then be generalized over the market framework, as follows. Demand and supply conditions produce an equilibrium at O . The policy makers' goal is then to stimulate the individual to consume a level of housing H^* , greater than H_0 . However, in order to ensure that these low income settlers have enough resources for housing as well as "other" consumption,

^{1/} This discussion ties back to the detailed discussion in Chapter 3.

Figure 8.1

8-15

policy makers impose another constraint. The households should spend no more than a certain percentage (say, $= .2$) of their monthly income on housing. Thus p^*h^* , or the amount spend on the desired minimum housing consumption should be less than a given percentage of the income level. The latter amount corresponds to a locus of prices and quantities along a curve which has the property that any point (p,h) on it implies equal expenditures. ^{1/}

Given X and Y , the relevant curve would define a target region of affordability. Assuming that the initial equilibrium defines consumption at H_0 in Figure 8.1, and a desired minimum level of housing consumption of H^* , the new equilibrium would have to lie in the shaded region abc in order to satisfy the affordability and minimum consumption criteria. To reach the target region, policy makers can attempt to shift either the demand or the supply curves, or both, although in the example, the supply curve must shift down at least as far as b . Demand for housing services can be boosted by measures that raise incomes, or the marginal propensity to consume housing. Any measure that raises incomes will tend to shift the affordability frontier as well. Supply can be shifted down and to the right by lowering the cost structure through various construction techniques, or simply by building or stimulating the building of more units.

^{1/} Each iso-expenditure curve would be a rectangular hyperbola. For each Y , there is a family of curves which vary with X . A changes in either X or Y would cause a shift to a different "affordability" curve.

At a very general level, then projects can be represented as attempts to induce shifts in market demand and supply curves, according to the services being offered by the components. Studies on housing consolidation and community participation can broadly be interpreted as statements on changes in determinants of the supply side of the market, whereas demand studies and specialized studies on employment generation tend to operate on the other side. Certain other components tend to operate on both sides.

Affordability concerns the location of the final equilibrium point in relation to its previous and desired positions. Cost-benefit analysis assesses the relative sizes of the producers' and consumers' surpluses before and after project implementation. It also examines whether the assumptions of perfect competition are valid, and the adjustments that can be made in relevant variables (prices and quantities) to account for market imperfections. It determines whether policy makers have subjective social goals that decide how costs and benefits are distributed.

On the supply side, projects' goals are to provide incentives for expansion of low-income housing services by increasing the land available for settlement; improving tenure arrangements, and providing infrastructure, services and innovative construction and financing techniques, etc. Prior to the Lusaka Conference most supply side analysis had concentrated on assessing mutual-help and self-help as construction methods. Mutual help had been used in constructing core units in El Salvador and some infrastructure in Lusaka. Self-help had

been intended as the primary construction method in all four projects. The success of these methods was being evaluated in terms of quality, time, and direct monetary costs with respect to the expectations of project designers and target populations, and possible alternative methods, such as contractor-built housing and informal processes of progressive development.

On the demand side, the projects' goals were more modest. In terms of the market framework, the project goal was not so much to shift the demand curve out as it was to raise recipients' welfare by moving down along the demand curve in supply shifts which lower the price of a unit of housing services through gains in efficiency and plot availability.

There are a number of ways in which operation on the demand side can contribute to desired results. First are direct income transfers, which were not incorporated into the projects under discussion. Second are income transfers contingent on participation in the project. These include payments for supplying labor to the project, and income transfers effected by providing project components at lowered prices. The former may occur either through private sector contracting and subcontracting or through mutual-help programs. In either case, the net income accreditation is: $Y = (w - oc) t$,

where: w = the wage rate paid per hour;
 oc = the hourly earnings foregone by participation; and
 t = the time involved.

To the extent that component price reductions stem from cost reductions, the response may be thought of as a change in the marginal propensity to consume (though this is strictly true only if the general price index remains unchanged). To the extent component prices are lowered still further, the resulting subsidy will produce changes both in real income and marginal propensity to consume.

With these several options, the choice to limit explicit demand stimulation largely to credit facilities and the provision of tenure hinges mainly on the concern to ensure projects' "replicability" (interpreted here in the shallow sense of recovering full costs). Subsidies or income transfers would restrict the local capacity to repeat projects, due to overall resource constraints. However, this choice also implies a greater reliance on the instruments affecting the supply curve to produce the desired level of housing consumption at a price affordable to the low-income population. The results will thus be more sensitive to supply side constraints, which had surfaced in problematic proportions, particularly in the African countries.

Credit financing is used in the projects to enable participants to afford initial capital outlays, and to enable families with variable incomes to keep up steady monthly repayment. In theory, it should be possible to calculate whether these provisions are at full cost (market value) or provide some subsidy. In practice this is not so easy to do. There are three general ways in which credit charges may be reduced or credit levels increased in a particular sub-market: 1) costs may be reduced by discovery of more efficient procedures or ways around

market "imperfections"; 2) credit in officially controlled channels may simply be diverted; or 3), through outright subsidies. Though the last is a clear cut case, the other two will almost always involve some element of subsidy as well, complicating the analysis.

The provision of tenure may occasion what seem to be non-monetary boosts in housing demand through the project which lower the implicit unit price of housing services for the owner-occupier by removing the "risk premium" associated with illegal settlement (a supply side phenomenon). Tenure also has an effect on the demand side, since the provision of land ownership may be a form of implicit subsidy if it is provided at less than the market value.

The project goal of replicability leaves little scope for subsidy. However, because of the trade-off between service levels and affordability, it became necessary to review some options.

Cross subsidy schemes were being tried in the Philippines. There commercial and industrial areas were being set aside so that the users of those areas could be charged higher rates, and the ensuing revenues applied to lowering the per-unit charges to residents.

Subsidized credit systems that do not require full repayment are a favored means of stimulating housing growth. A conventional credit system may become tacitly a subsidized one in the event of frequent defaults.

An implicit land subsidy has been granted in Zambia, where land has been nationalized. It may also play an important role in the

Philippines, where the project is centrally located in Manila and it appears is being made available at less than market value.

Achieving less than full cost recovery, like subsidized credit, is something that may be planned, or simply tolerated because of the high social cost of cracking down on a massively defaulting population.

In the simple demand/supply framework mentioned earlier, the concept of affordability corresponds to the location of final equilibrium position in relation to its previous or desired position. A non-affordable solution may result from policy makers' inability to affect the demand and supply curves as desired due to institutional or resource constraints. It may also result from miscalculations of the curves' locations and determinants by policy makers (in setting a minimum amount of desired housing consumption) or by target populations (in estimating their own ability to generate income and savings over a specific period). A sites and services scheme will be deemed unaffordable if it does not succeed in attracting participants whose incomes are low enough to place them among the target population. Even if the target population is reached, there may be indications that the project is not really affordable. Participants may drop out because of project-related pressures on their incomes, or stay in the project without consolidating their houses reasonably quickly, or they may devote too large a proportion of their income to housing, to the detriment of other needs. They may also simply default on their service charges or loan repayments.

An upgrading scheme must be considered unaffordable if residents move from the project area in substantial numbers. Even if they do not leave, some families may be overextending themselves to stay.

Initial Findings

Initial findings of the Evaluation indicated that the basic sites and services and upgrading approaches seemed to be working, but not without a number of flaws. Affordability results were so far mixed. It became clear, though, that the search for solutions to affordability problems would have to extend beyond the narrow limits of the projects themselves. Only probing into the basic workings of the housing and labor markets and full cost approaches would yield the requisite understanding of the matter.

The following are some of the evaluation findings indicating the validity of the sites and services and upgrading approaches. The general finding of a study of the mutual-help scheme in Sonsonate, El Salvador was that mutual-help and self-help labor can produce dwellings of approximately the same structural quality as those built by contractors. ^{1/} The same study revealed that mutual-help is also an efficient means of housing construction. With the wage paid roughly equivalent to average opportunity cost, total construction costs under the mutual-help scheme were ten percent less than those for similar

1/ FSDVM Report #12. [Full]

units built by contractors. ^{1/} Two studies in El Salvador and one in Senegal indicated further that self-help methods will also work well. ^{2/} The latter study, of households in the control group areas, established that families with incomes in the target range were able and content to use both self-help and progressive development methods, with the construction process typically stretching over two years. Not being project beneficiaries, these families (84%) constructed their homes without formal loans, relying on transfers from the extended family and on rent from the addition of rooms. These were the first findings to suggest the importance of income transfers in providing families with needed resources for housing improvement, providing an indication that, in evaluation, it is an ill wind indeed which blows no good. It is precisely because there was a rigorous multi-country evaluation being conducted, that findings emanating from an investigation of control groups could be assimilated as harbingers of project participants' behavior.

The availability of construction loans seemed also to be an effective way of spreading housing expenditures over time, and thus to be playing its planned role of making project housing accessible to

^{1/} This finding must be qualified. The cost difference apparently resulted from the fact that commercial contractors have higher administrative overhead due to various contributions for social benefits. The favorable comparison is based on private costs and benefits only and results from a subsidy "extracted" from government. Still, even on a social cost-benefit calculation the comparison is not unfavorable.

^{2/} Studies of the San Jose del Piró (FSDVM Report #) and Usulután (FSDVM Report #11) sub-projects; study of the control group areas in Guediawaye and Grand Yoff (Bureau d'Evaluation Report #4).

families with variable incomes. 1/ The use of credit for construction had been widespread in the project areas, exceeding that among households in non-project areas surveyed. 2/

Finally, experience in all project contexts had tended to confirm the importance of enlisting the participation, or at least the cooperation of the communities in project execution. In addition, findings indicated that working with existing community organizations had been much more effective than creating new ones. 3/

However, there were plenty of problems to temper the basic success story. A fundamental and general set of problems arose from delays in the implementation of nearly all the projects. Such delays have emerged as a difficulty with virtually all urban projects, to a lesser extent in Bank-financed projects than in many others. The stalled Senegal project showed that the time required for project implementation, and specifically house construction, can have a serious

1/ FSDVM Report #

2/ In the Dakar sites and services area, only 47% of the household heads who had started construction were taking advantage of loans. However, almost all successful applicants had applied for loans to purchase their plots. In the control group areas of Guediawaye and Grand Yoff, 84% had constructed their homes without the help of loans. In Zambia, almost all participants had applied for building materials loans. However, most have found the loans inadequate for building even a core unit.

3/ In Zambia, the political party (the UNIP) was instrumental in mobilizing the community to participate in mutual-help construction of infrastructure (LHPET Report #3 and #13). The entire Philippines reblocking scheme was carried out with the active support of "baranguay" leaders (each of whom represented about 470 households in the target areas), (RAD Report #1). The problem of community organization was aggravated in the Senegal sites and services scheme, where it was proving difficult to organize applicants drawn from a variety of locations who could not yet come together on the site.

financial impact on residents. If they do not build temporary shelters as the construction of more permanent dwellings proceeds, residents may be forced to pay charges for both their old and new units for an extended period. Because of inflation, a long construction period also implies greater nominal and real costs of materials and labor. Evaluation studies have tended to confirm, and in some cases quantify, these observations. ^{1/}

There were several specific findings in this sphere. In El Salvador, the process of mutual-help construction had taken an average of ten weeks longer than the thirty-six week period allocated by project designers. ^{2/} Difficulties included a misallocation of labor and a general lack of technical expertise in the formation of work groups; inadequate instruction in building techniques by technical assistants; absenteeism ranging up to about 15%, concentrated among the more highly skilled workers who incurred greater opportunity costs by participating in weekend work; and leadership problems. The work groups' efficiency depended on leaders' organizational abilities and on inputs from Fundacion social workers.

The evaluations confirmed that construction standards are directly related to the time for housing consolidation, and to its final cost. In Senegal, the very slow rate of house consolidation appeared to be a direct result of structural requirements set by project management that were inconsistent with the incomes of portions of the target

^{1/} Bureau d'Evaluation Report # .
^{2/} FSDVM Report #12.

population. ^{1/} In Zambia, there was evidence that technical assistance provided to residents was biasing construction towards higher quality building materials (cement, rather than sun-dried, bricks), and towards higher final costs with only marginal returns in structural soundness.^{2/} Unduly rigorous codes and restrictions regarding the quality and conditions of legally habitable dwellings have also added greatly to the duration and cost of construction. In the Senegal project, which had been appraised in 1972, of 1,763 resident-allocated plots, only 187 households conforming to project criteria had begun building as of fall, 1977. The main reasons for construction delays were exigent legal procedures that led to waiting periods of from 7 to 18 months between allocation and ownership.

There were other findings that revealed obstacles to efficient project operations. Resource constraints hindered house construction, particularly in Africa. In Zambia, a time budget study showed that families engaged in self-help construction spent nearly twice as much time obtaining materials as working on construction during the first 2 to 3 months following resettlement. ^{3/} Difficulty in obtaining materials was also cited in studies carried out in Senegal. ^{4/}

^{1/} The model desing layouts proposed by OHLM, the most popular of which then cost about US \$6000, necessitated long savings periods, at best. At worst, they led to long periods of stasis while successful applicants awaited a compatible loan program (Bureau d'Evaluation Report #7.)

^{2/} LHPET Report #12.

^{3/} LHPET Report # .

^{4/} Bureau d'Evaluation Report # .

It also emerged that setting appropriate limits to loan programs would be an exacting task. The Senegal project had no loan program initially, and this apparently contributed to delays. This initial oversight was soon corrected. The striking fact is that participants' expectations about their houses had to be adjusted downwards before it would be possible to design a loan program that would permit the houses to be built and the residents to repay. ^{1/} The Zambian evaluation found also that the loan ceiling had been set too low to permit very many residents to build even the intended initial two-room house. ^{2/} Initial repayment performance had also been poor in the two African countries. The poor repayment rate may in part have been a result of unrealistic expectations, some families having overestimated their future incomes and repayment capacities. What is more likely is that high default rates should be expected among populations unaccustomed to paying for services or repaying loans, particularly where, as in Zambia, there has been a long tradition of subsidized housing at all levels. ^{3/}

The importance of developing "contextual" information and understanding the whole housing market was also highlighted by the project experience in El Salvador. Although project-built housing provides a higher overall level of services per site, the quality of the dwelling units themselves may not differ markedly from that of some

^{1/} Bureau d'Evaluation Report # .

^{2/} LHPET Report # .

^{3/} The degree of subsidy varies, as such programs treated housing partly as a vehicle for paying wages in kind.

other types provided by the informal, progressive development market. This appeared to be the case with the colonias ilegales in Usulután, El Salvador, 1/ as well as the control areas in Dakar. 2/ This factor may mean that project housing is not always seen to be competitive with other housing options available to the target population. For example, in Usulután, only 18% of those sampled in the illegal subdivisions were willing to participate in the project. Most who did apply were from mesones (or rented rooms), and even the few applicants from colonias ilegales tended to be renters there. Among the reasons for this was the fact that unit prices of the FSDVM project tended to be higher. In effect, the location of the demand curve had been miscalculated; that is, policy makers' expectations of the level of housing services individuals within the target income range would be willing to consume at given prices were too high. Government codes also set housing standards too high to reflect the preferences of the population. People wanted more land, fewer services, and lower costs those standards implied, 3/ as was the case in Dakar as well.

Studies in all countries also confirmed the need to develop contextual information on labor markets. Against expectations, it was found that project beneficiaries tended to prefer hiring commercial contractors or individual laborers rather than performing the house construction themselves. In Senegal almost all those who had started building had hired contractors, claiming lack of time as the main

1/ FSDVM Report #11.

2/ Bureau d'Evaluation Report #4.

3/ FSDVM Report #11.

reason. In one of the Zambian resettlement areas (the John Howard area), 80% of the participants used contract labor. These findings were taken to indicate that families' own labor involved high opportunity costs, either in terms of earnings or of foregone leisure, and/or that the technical assistance provided by the project was not adequate to make people feel confident of being able to build for themselves.

It is worth mentioning again in context that the first studies in El Salvador showed that the opportunity costs of participants enrolled in mutual-help programs were much higher than had been assumed by project designers; furthermore, they were quite variable, so that whilst some families could participate at small sacrifice, many were participating despite a net loss of income or were dropping out to avoid this. ^{1/} The finding exemplifies an infrequently recognized value of evaluation studies. Mutual help was organized for weekends, reflecting a developed-country, formal sector mentality on the part of project designers. It was assumed implicitly that opportunity costs would be lowest on weekends. For some this was true, but for many others, especially those in informal sector activities where weekends are peak earning periods, the opposite was true. Project design should reflect these realities. Though the case of self help is different in that families can organize work for periods when opportunity costs are lowest, it is still true that for some opportunity costs are never low enough.

1/ FSDVM Report #12.

In El Salvador, it was found that the full cost of housing to beneficiaries exceeded the monthly charges levied by the implementing agencies to cover plot and core structure by much more than the presumed . 1/ Based on information gathered in early stages of the project in Usulután, it was estimated that, over the first three years of the project, the total payments covering building materials, labor, and installation costs for a three-room house would be twice the charges levied by the Fundacion. This kind of discrepancy, and those relating to opportunity costs, indicated that much more attention would have to be given to full and accurate accounting of costs as a feature of all projects.

The findings already summarized appeared to some to bode ill for the affordability of projects. The early evidence on affordability showed that the level of services offered was unattainable to about the poorest 25% of the population in El Salvador. The lowest two income deciles of the population were under-represented. And, if incomes had been faithfully recorded, those in the projects were overreaching themselves in terms of the affordability criteria spelled out earlier. To reach these people, without causing them to overextend themselves, management must decide whether to lower the service level in sites and services areas and concentrate more on upgrading (where affordability evidence appeared more promising), or to boost the demand curve by the use of subsidies.

1/ FSDVM Report #13.

In Senegal, the available evidence was more tentative, but interesting. Most of those being selected for the project appeared to have incomes within the prescribed range. However, the distribution of participants' incomes within the prescribed range tended to cluster towards the upper end of the scale, and the lower incomes in the range appeared inadequate to cover the full costs of the housing services being offered. Even though the incomes of 3,500 selected households were biased towards the upper end of the prescribed range, households were generally still having problems meeting the costs of construction.

In addition, there appeared to be biases in the selection process pertaining to factors other than income. Initial findings indicated modest biases against large families and those with informal sector employment. The most significant bias was based on the sex of the household head: 16% of applicants were women, whereas only 10% of recipients were. It appeared that this result was a concomitant of the income bias, in that female household heads earned on average \$10 less per month than their male counterparts.

The Zambia sites and services sub-projects seemed to be producing results that were reasonably close to the mark. However, the project seemed to be developing a relatively high dropout rate. Only 25% of those who applied and were selected had declared an intention to move to the site by paying a deposit. ^{1/} In the resettlement areas, the dropout rate among people whose structures were demolished was about 31%. The Lusaka Housing Project Evaluation Team hypothesized different

^{1/} LHPET Report #11.

criteria for affordability (based on basic needs concepts) to try to account for these findings. Its suggestions were not accepted at the Lusaka Conference and were in fact roundly criticized.

The information that had been gathered by that time on upgrading schemes in Zambia and the Philippines was only preliminary. Because the second applications of the longitudinal surveys had yet to be analysed, there were no obvious indications that the projects were not affordable. Because the upgrading schemes require considerably lower initial capital outlays (of money or labor), participation in them was not expected to strain the budgets of residents.

Reformulation of Hypotheses

Initially, the evaluation models considered within the World Bank concentrated on goal-attainment and impact evaluation, to address such questions as how much housing and employment resulted from a project, and whether the results were different from project objectives and expectations. With its analysis of the initial evaluation findings, the Lusaka Conference began to shift activities toward process evaluation, focusing on measuring and accounting for various types of behavior motivated by a project. It was this pivotal change of orientation that necessitated the reformulation of many hypotheses. A few examples will illustrate the process that occurred.

The early findings convinced many of those concerned that progressive development and self help were distinct features of the housing development models the Bank was fostering and supporting. Until this time, the two terms had been used almost interchangeably. Then the

early research results included cases both of self-help construction completed in a short period and phased construction of houses relying mainly on hired labor.

"Normal" behavior in the informal housing market might consist of the quick construction of a shelter using "temporary" materials, followed by its phased extension and improvement using durable materials. The evaluation at first showed that certain projects seemed to conform to this pattern: San Jose del Pino, an early project in El Salvador; Grand Yoff and Guediawaye, the control areas in Senegal; and the George overspill areas in Zambia. In the Dakar project on the other hand, and in most of the Lusaka sites and services areas, construction was either proceeding rapidly or not at all. These observations led to a series of questions: Did the projects tend to discourage (or fail to encourage) progressive development in some ways? How did these tendencies occur? Were they ever desirable? On the basis of such questions hypotheses were reformulated (or formulated for the first time) involving the effects on construction of the distance participants were required to move, the availability of "temporary" materials, the cost of their transportation, the availability of credit, etc.

Another key question was the extent to which project participants' unexpected reliance on hired labor reflected "errors" or inflexibility in project design. Reformulated hypotheses in this area focused on relationships between the proportion of self-help labor used (dependent variable) and families' skill endowments and opportunity costs (independent variables).

Several research initiatives were taken at this time. First, although it was acknowledged that extensive use of hired labor was not undesirable, attention was turned to certain instances in which project design tended unnecessarily to restrict participants' options to use their own labor. Findings from the first study of mutual help in El Salvador suggested numerous ways in which the mutual-help component might be redesigned for greater efficiency and greater equity among families, as it was found that some pressures on mutual help correlated with certain family characteristics.

A second decision was to develop a model and assess families' decisions to devote their own labor to house building, rather than offering more of their labor in the market using their increased earnings to hire commercial labor. The same model would be used to assess associated variations in total housing demand, total labor demand, and substitution between labor and materials.

A third set of research decisions centered on the evidence that the projects might be biased against some of the economically weakest elements of the target group: large families, those with low incomes, the self-employed, households headed by females or single persons, etc. It became important to learn whether the projects were systematically excluding such sub-groups. Did they perceive the project as unaffordable, inconvenient, or an otherwise undesirable dwelling option?

A final set of observations hinged on the finding that families that had previously been renters were under-represented in the

projects. Such families were applying for and entering sites and services projects in disproportionately small numbers. This finding called attention to the fact that the projects had been concerned principally with owners and ownership, and very little with the needs of short-term renters. In some instances rental arrangements were proscribed or severely restricted. In short, there might be certain groups of "natural renters" whose participation was restricted because the projects lacked design components geared to their needs. Such "oversight" reduced the opportunities for renters to participate in the projects and so improve their housing situation. It also limited the options of owners and potential owners who might consume a higher level of housing services if they could rent part of their own dwellings to others. Accordingly several hypotheses in this area were reformulated and appropriate tests incorporated into the work program.

The research designs and evaluation programs were altered along the lines sketched above during the subsequent years. Not every question that had arisen could be addressed, nor could every new understanding be acted upon; however, several major components, outlined below, were added to the program. These adjustments to the research program could be thought of partly as improvements and partly as extensions. Improvements resulted from learning experiences that stimulated new hypotheses or the reformulation of old ones. Extensions involved pushing the analyses into new areas or to a new level of complexity or detail. Some of the extensions had been foreseen and planned for, others had not. At this point, the case was starting to be

made for the input of research resources beyond those directly available to the evaluation program, something that had been anticipated at the beginning of the program.

Supply and Demand Analysis. As discussed earlier, affordability results had been expected to depend on how the projects affected the demand and supply sides of the housing market. Early results confirmed this expectation and pointed to some areas for further study. A large portion of future evaluation research would concentrate on assessing the determinants of the supply and demand curves, in the hope of gaining insights into the workings of the low-income housing markets at the micro level. To some extent, this would be accomplished through separate studies that would isolate particular effects.

The early findings on self-help and contract labor use did not establish whether there was a real basis for preferring one method over the other. But these findings did clarify evaluators' thinking on two points. First, self-help per se (i.e., use of own labor) should never have been regarded as an end, but rather as a means to be facilitated. Second, we had not been able so far to investigate in enough detail the relative implications of self-help methods for specific purposes in well-defined market circumstances. Initially, this implied more detailed study, and in the long term it implied more options in project design. Thus, for example, the evaluation would examine technical assistance programs for their efficiency in fostering requisite skills; consider various project design features in terms of their avoidance of

high opportunity costs; and assess the relative effects of materials shortages on different construction methods. ^{1/}

The evaluations would be at a disadvantage in assessing the determinants of demand because the projects lacked instruments for the stimulation of demand. Still, this was regarded as an important area of investigation, and some progress had been made in conducting "demand studies" in El Salvador. These were studies launched when sub-projects in particular cities seemed to be encountering sluggish demand. The studies compared the project options with the broader range of options available in the market--legal, quasi-legal, and illegal--and the results were illuminating. It was decided that such studies should be pursued, even if they led to no broader range of options within the projects themselves, and eventually extended to the estimation of price and income elasticities of demand.

Cost-Benefit Analysis. Of the possible ways of integrating analyses of supply and demand, cost-benefit analysis was the first adopted. The motivation here was to combine research into lower-cost options for reaching a broader spectrum of the poor population with the comparison of project options and market options. Initially, investigation of the lower range of the scale would be limited to illegal or quasi-legal options provided by the informal market to discover what lowered level of housing services would enjoy a vigorous demand. The benefits and costs would be assessed from the points of

^{1/} In the event, studies along these particular lines have not so far been undertaken to the extent contemplated at the end of 1977.

view of both the individual household and the society as a whole. In addition to assessing the overall ratios for projects, the cost-benefit framework would be used to review the appropriateness of project components for achieving their stated goals.

The Economics of Self-help Housing. Questions dealing with the use of self-help methods cannot be considered in isolation. To some extent, the use or stimulation of self-help methods in the projects may be viewed as a temporary means of getting around certain market imperfections, though it is also true that self-help methods constitute a part of the market solution to housing problems. (Studies have shown that this is true even among the highest income populations in the highest income countries.) Relationships with contract labor are sometimes competitive, sometimes complementary, and the key questions are not about the relative proportions of the two methods' uses, but whether opportunities for cost-savings by one method or the other are being overlooked. Similar considerations should apply in the examination of trade-offs between self-help and mutual-help methods, and between labor and materials. To do this, a formal model was necessary wherein the microeconomic choices of the consumer, as a potential producer of what is to be consumed (housing), can be examined in the framework of optimal time allocation. ^{1/} We would thus hope to learn more about the scope for mutual help and self help in various circumstances. Given information from these analyses, the total costs

^{1/} Since the problem is analogous to that of the peasant farmer considering in what proportions to produce for his own consumption and for the market, an analogous model was adopted.

of the various techniques would be compared to obtain preliminary indication of the location and behavior of the supply curve, and the relationships between the wage rate, labor supply and demand, and housing consumption.

Valuation and Hedonic Analysis. The development of reliable means of valuing squatter, improved, and sites and services housing had been assigned a high priority from the beginning of the evaluation program. Project and program planning required general measures to assess the costs imposed by the destruction of squatter housing, as well as the values created by the projects, whereas the statistical relationship between value and particular housing characteristics would permit derivation of the shadow prices of these characteristics, or project components.

Little was accomplished in this area during the first two years simply because, except in El Salvador, houses were not yet being built or improved. However, enough preliminary work had been done to make possible agreement among researchers that the questions concerning how much houses are worth and what determines their value would be treated simultaneously, using various systems of appraisal and hedonic price techniques.

Assessment of the Program: Lessons Learned

Developments initiated in 1978 as a result of the first two years' work and conferences were sustained and extended during the years that followed. Those years have seen a maturation of ideas, models, and processes, but little further change in the basic directions of the research.

In 1979, an administrative decision was made that had important consequences for the evaluation program. Up to that time, about one additional field evaluation per year had been coming under the supervision of the Bank evaluation staff in charge of the program, and this trend has continued. ^{1/} What changed in 1979 was the evaluation staff's level of involvement in the supervision effort. In the initial three countries, in the Philippines, and in Indonesia, the degree of involvement had been high, and to good effect. However, this effort was beginning to claim an excessive share of the resources available for urban research. The 1979 decision to streamline drastically any new supervision efforts and to concentrate resources on completing the pilot program was an exceedingly fortunate one. Without it, the pilot program would probably never have been successfully concluded, the research output would have been substantially less voluminous, and this book would certainly never have seen the light of day.

The story of the Evaluation is essentially a happy one; however, not everything went as it should have. As there is more to be

^{1/} Starting with the first Philippines Project in 1977, Indonesia was added in 1978, Colombia in 1979, Kenya and India in 1980, and Bolivia in 1981.

learned from mistakes than from successes that come to be taken for granted, this chapter section discusses the principal shortcomings of the approach taken.

Shortcomings of the Approach: Harsh Encounters with Reality

Errors on the Demand Side. Certain errors in the planning of the Evaluation compromised its results. at least for a time The errors fall into three categories. All had their roots in the organizational structure of the World Bank which has tended to separate operational and research responsibilities, and to deprive operational departments of reasonable resources for evaluation research.

The first set of problems emerged from the Bank's tendency to define what was needed for guidance in terms of monitoring and evaluation. Neither of these terms has been properly defined by the institution even to this day.. With respect to the program under discussion, monitoring tended to be viewed as a "quick and dirty" version of evaluation, and the two concerns tended to be referred to as big M (or E) and little m (or e). Then the bureaucratic decision was made to assign big M (or E) activities to research staff and little m (or e) activities to projects staff. Thereafter the activities proceeded more or less independently, with unfortunate consequences.

Monitoring and evaluation activities are not different levels of the same thing, nor should they try to be. Yet that is how they came to be regarded in the World Bank, and still are, by some. As described above, evaluation deals with a process of determining how inputs are

converted into outputs and these in turn into effects and impacts. Monitoring, on the other hand, is a technique for gathering information on a steady basis about on-going processes.

Not so in the Bank of the early 1970s: Evaluation became the name for doing what this book describes, while Monitoring was taken to mean doing the same thing on a smaller scale. This had all kinds of undesirable consequences, not least of which was that research personnel responsible for evaluation initially neglected short-term effects, assuming that those were being dealt with by others. This incorrect assumption had deleterious effects on the substance of the program in its first years.

Then, because evaluation became the responsibility of a research division (while monitoring was supposedly being handled by another unit close to operations), the evaluation program tended to become overly rigorous and circumspective. ^{1/} If this was all that had happened, the damage might have been fairly limited. Unfortunately, there was a lasting effect: the evaluation program has yet to devote enough attention to creating means by which findings of immediate operational use could be quickly disseminated. ^{2/}

^{1/} A better balance would certainly have been desirable, though the assignment of evaluation to an operational unit would not have been. The emphasis on the "quick" would have made the process impossibly "dirty", as has been the case to date with all evaluation activities excessively concerned with speed.

^{2/} The outcome might have been different in the available technology had been different. Now that the computer graphics and mapping are available within the Bank, it is easier to conceive of solutions to the program's past shortcomings.

In retrospect, the initial research design was also overly ambitious, with too many topics for investigation and very high standards for the specification of findings. Though they were not blameless, it was not the evaluators who were primarily at fault. It was operational personnel who had endless lists of questions they wanted investigated, with very little sense of the trade-offs or opportunity costs involved. The excess of research topics actually reflected a lack of clarity as to the subjects worthy of investigation. This unclarity reflected a further uncertainty on the part of operational staffs as to the range of objectives possible in the early multi-faceted projects. The evaluations were too ambitious in their assumptions about what could be credibly learned, in the areas of health and education, for example. Finally, the issue of the relative balance between the qualitative and quantitative tools of evaluation was never raised during the early going. Research and evaluation staff tended to think automatically in terms of the experimental or quasi-experimental design. Operational staff did not have an evaluation model in mind, only questions to which they urgently desired answers. Fortunately, once these initial errors and their attendant difficulties came to light, the appropriate compensatory adjustments in the program were readily made.

Errors on the Supply Side. Faulty judgments were made at the outset regarding the capacities of host countries to support the conducting of evaluation activities. Planning did not take account of either the scarcity of adequately trained personnel or the severely

limited data and data processing capabilities in the countries chosen as project locations.

These oversights led to errors in efforts to tailor the evaluation model to individual country circumstances. There was a definite failure at the outset to recognize the crucial differences between situations in the African countries and those in El Salvador and the Philippines. Evaluation models suited to the latter regions were applied essentially without alteration to the very different situations in Senegal and Zambia.

The choice of Senegal as a host country was unavoidable, given the genesis of the program, and it could even be argued that the evaluation model was suited to the country, particularly in view of the setting -- a large continuous project. However, there is no doubt that Zambia was a questionable choice, as some operational staff argued vigorously at the time. Certainly, once it had been chosen, even a cursory assessment of its capacities -- data, data generating capabilities, human resources in the social sciences, computer capabilities and institutions -- would have dictated a much simpler evaluation model.

One result of this error was that the Zambian evaluation was not cost effective. ^{1/} It lagged behind the others at all stages. Recruitment was slow, and in the end inadequate. The staff recruited

^{1/} Neither was the Senegalese one, but primarily for different reasons: the planned evaluation probably could have executed had the project been implemented, but it was not within the time allotted to the Evaluation.

were still slower in developing the research design. Then implementation of the evaluation (more than of the project) lagged badly. Even in spite of these lags, the evaluation staff clung slavishly to the original evaluation model, and failed, until quite late, to make the kinds of adjustments that were readily made in the other countries.

The lesson of these experiences was that evaluations must be tailored to particular country circumstances, as well as to other factors.

A relatively neglected area in all the evaluation was "institutional" questions. The neglect took the form of evasion of questions such as how staff were to be selected and trained, and how the established evaluation operations were to be administrated in the largely unfavorable circumstances. The importance of the latter question had been recognized very early in the planning stage. However, we failed to keep it in the forefront of our thinking, and several years' opportunity to improve the supply situation in Senegal and Zambia were lost. ^{1/}

^{1/} Here, as in other instances, the essential error lay and still lies at the top of the Bank. There was no suggestion of funds being made available -- even indirectly if not directly -- for such purposes.

Except for minor errors of judgment which caused some loss of credibility among operations personnel, 1/ the remaining areas of disappointment were technical. In the case of Zambia, the failure to adapt the evaluation model to the actual existing capacities for reliable data collection, analysis, and reporting, resulted in considerable "data overhang". A considerable body of data was collected and not well utilized, either because it was deemed unreliable or deficient, or because of insufficient analytical capacity.

Such problems did not develop in El Salvador or the Philippines. 2/ However, even in these countries, the evolving needs of the evaluation were such that the initial instruments did not supply all the needed measurements. In fact, this outcome is hardly ever averted in first research endeavors of any kind. No matter how careful the planning, there is an inevitable lack of fit between initial data generated and the model or models eventually developed. 3/

Finally, during the technical (planning) conference in January, 1976, a decision was made to tailor each evaluation to

1/ A case in point is the attempt made in Senegal to conserve funds and achieve a sample census by using the project application form as a baseline survey documents. A potentially good idea was mismanaged on the ground, caused delays, and contributed to early alienation of operational staff responsible for the project.

2/ In Senegal the problem was of a different nature: much of the intended research was never carried out because project delays meant that a population of sufficient size to make data collection possible was never available.

3/ I am indebted for this observation to Greg Ingram, who cites similar results from his experience working the NEBR urban model. See Ingram, Gregory K., "Discussion of Housing Behavior and the Experimental Housing Program: What Have We Learned?" - Paper presented for the NEBR Conference on Social Experimentation, March 5-7, 1981, Hilton Head, S. C.

particular projects and country circumstances. This decision consciously curtailed significantly the opportunities to perform comparative analyses using data from the four countries, and it was not an error. In fact, it still seems justified. It is mentioned here as a good illustration of the kind of trade-off that must inevitably be made in such exercises. Not everything highly desirable from the research point of view can be accomplished: many worthwhile analyses cannot be executed and some which can be should not be for cost effective reasons. Many things which are attempted will go unachieved in the end, either as sacrifices to the inevitable imbalance between objectives and means, or because unpredictable developments rule them out. At the same time, some analyses will be completed which, in retrospect, will appear to be of little use. The lesson is that selectivity in the setting of objectives is essential, but foresight is never perfect.

Assessment of the Special Evaluation Program: Achievements and Limitations

In 1972, there were no proven methodological approaches to the type of evaluation undertaken within the World Bank. From the outset, the process involved a "learning by doing" approach, as did the introduction of urban lending itself. The following pages evaluate the study against key criteria, in order to identify its achievements and limitations, and to derive lessons for future efforts.

Substance and Coverage of Subjects: The major substantive focus of the special study was on the behavior of households under

project conditions, including the rate of housing construction and the factors that influenced that process. Specific areas that received detailed attention included:

- * the selection of households to participate in a project;
- * the rate of construction;
- * the organization of construction, through self-help, contracted labor, or some combination of the two.
- * the quality, cost, value, and physical characteristics of the housing constructed;
- * the affordability of the project at the household level, covering both payment for the serviced site and repayment of the construction loan;
- * household performance in meeting mortgage obligations, i.e., cost recovery;
- * impacts of the projects on socio-economic conditions of participant households through employment generation and income growth;
- * impacts of the project on city-wide housing stock; and
- * impacts of the projects on urban policies.

In each of these areas, the special evaluation program was able to reach rigorously supported conclusions, the most important of which are presented in Part II.

These were the subjects most discussed at the outset of the Bank's urban lending. The greatest degree of uncertainty pertained to

the facts at the household level. Would project households invest in housing under conditions of secure tenure and affordable infrastructure? On the basis of the special study and of other project studies, the answer is certainly positive. If anything, households required less stimulus than anticipated to conduct housing investment on a substantial scale. The degree of uncertainty early on, however, made the special study very much a product of the formative years of Bank urban lending.

Still, it has not been possible to measure or otherwise find out about all the impacts expected from the projects. The greatest factor working against this effort has been time. In part because of delays in project implementation, not enough time has elapsed for some impacts to have occurred. ^{1/} Looking ahead, even the passage of time will not permit some impacts -- notably those related to changes in health and education -- to be measured within the evaluation format. In the case of education, the gestation process for important impacts is simply too long and indirect. Thus, the most that can be hoped for from evaluation in this area is that it will collect and sort data that will facilitate future research into such matters. Such a limited effort must still be considered an important objective of integrated evaluation programs.

At the beginning of the DEDRB study there was particular interest in the question of the projects' health impacts on

^{1/} Delays averaged 18 months per project; several years in the case of the Senegal project. These lags compressed the data gathering period for the five year program to less than three years.

participating households. So, despite virtually unanimous advice against it, on the part of several experts consulted at the planning stage, much effort went into planning research on health-related issues, and extensive questions on health were included in the questionnaires for the four host countries, all to very little avail. Getting reliable information on health conditions and forming accurate estimates of projects' impacts on health changes require special research designs and measurement techniques. ^{1/} Extended discussions with health experts also revealed a lack of consensus as to the appropriate methodology for the evaluation of health impacts in complex projects. Research proposals for associated special health studies were prepared for El Salvador, the Philippines, and Senegal, with budgets ranging up to \$250,000 and more. The El Salvador proposal was approved by the Bank's Research Committee, but it was not possible to find an executing agency for the study that could have carried it out successfully and within budget.

The result has been that health issues have not been addressed so far in the evaluation program. Furthermore, it is doubtful whether they should be. Effective evaluation of health effects and impacts seems to require a huge research apparatus, transcending the scope of these projects, and making costs prohibitive.

^{1/} Even if successful measurements could be made of changes in health status and of the functioning of various health-related project components, attribution of specific changes in health status to particular components would be extremely difficult, if not impossible. These difficulties were acknowledged, but underestimated, at the outset.

Estimation of cost-benefit ratios and cost effectiveness studies of particular project components, especially those related to new or enhanced housing quality and value, have generally yielded positive results. Contributions have also been made to strengthening national housing authorities' capacities to meet their respective individual country shelter needs and to the tailoring of evaluation research designs to the needs of management. These points are discussed further in the final chapter.

It is worth noting the efforts made by evaluation research staff to clarify housing approaches and to develop an operationally-oriented basic needs theory appropriate to them, although some of the work is still in its initial stages. The following is a partial list of subjects on which conceptual and theoretical progress has been made:

- (1) the formulation of a more accurate definition of self-help;
- (2) the development of a three-factor model to assess family decisions with respect to labor used in the housing construction process;
- (3) the measurement of income and expenditures, and the discovery of the role of transfers;
- (4) the development of a "contractarian" model to explain transfer behavior;
- (5) improved project cost concepts ;
- (6) the theoretical and empirical comparison of informal housing to informal employment phenomena;
- (7) improved ways of viewing relationships between housing and poverty alleviation objectives - and sites and services and squatter upgrading strategies - growing out of the study of project beneficiaries;
- (8) the application of hedonic analysis to questions of housing and component value in the projects under study; and
- (9) the

articulation of a framework for assessing component-level and project-level efficiency trade-offs. ^{1/} These studies, and the tracking efforts undertaken by Bank and country operations officers, managers, and the evaluation units, appear to make the overall five-year IBRD-IDRC evaluation effort as credible as any that can be found in the burgeoning literature on project evaluation research carried out in recent years.

Operational Relevance: A major aim of the research design was for the outputs to be operationally useful to the project agencies in the four host countries. Though there were shortcomings, as discussed above, this objective was nevertheless achieved in varying degrees for each of the four projects, and in a broader sense for the entire Bank urban lending program. The assumptions tested are operationally significant for all the shelter projects with which the Bank has been involved. At least 20 operationally useful reports were produced by Bank evaluation staff - and more by the field teams in each country during the five year period. In each country the teams were able also to provide pertinent data and analyses to assist project managers in overcoming obstacles to implementation.

Although the teams were able to provide short-term conclusions fairly rapidly, the time and methods required for rigorous conclusions concerning impacts necessarily caused delays in the dissemination of

^{1/} In addition, work on the investment demand for housing, the tenure choice decision, and credit demand is now proceeding as part of the research program. See "Housing Demand and Housing Finance in Developing Countries," a research proposal submitted by DEDRB and approved by the World Bank Research Committee as Research Project 672-46. (Refer to Research News.)

findings in the agencies and within the Bank. The delays aroused impatience on the part of the intended audiences for research information and as a result, and somewhat misleadingly, raised questions as to the operational relevance of evaluation research to administration. ^{1/} It should by now be apparent to everyone that there are inevitably costs in time associated with the measurement of impacts over time. The same audiences that required methodological rigor were those who were most impatient with the time required to produce valid results. In the future, reasonable expectations on all sides might be established by addressing at the outset the trade-off between methodological rigor and the pace of evaluation.

A leading practitioner of evaluation has written that "utilization would probably be most widely accepted as the single most important criteria [sic] of the success of evaluation research." ^{2/} By this standard, the pilot program reviewed here scores well, whereas, according to the World Bank's Operations Evaluation Department, other monitoring and evaluation programs do not.

Methodological Validity and Research Design. A rigorous review of the research design and the study's findings demonstrates that the DEDRB effort was methodologically valid. As described in greater detail above the DEDRB study sought to compare the behavior of households within project areas with that of others living in other

^{1/} The tensions between operational and research staff, which were always present, and at times considerable, were nevertheless well managed throughout this exercise.

^{2/} Thomas D. Cook, Evaluation Studies Review Annual , (Vol. 3), Sage, Beverly Hills, CA. 1978.

neighborhoods, to measure project impacts. Since these impacts occurred over time, it was necessary to establish samples of project and non-project households, and to follow them through project implementation. While it proved difficult to maintain appropriate samples outside of project areas because of household mobility and the lack of communities with socio-economic profiles closely matching project populations, a number of statistical techniques were developed to ensure validity.

Another difficult issue concerned the representativeness of the samples when results had to be generalized to the total low-income population of the city in order to ascertain the city-wide impacts of projects. In most cases the problem was avoided by restricting conclusions to specific sub-groups, and caution was used when generalizing beyond those groups. Maintaining the methodological rigor of the study was also made difficult by the numerous uncertainties regarding the pace of project implementation and local research capability. The research design had to be adjusted frequently to take account of changes in the implementation schedules of specific projects.

It is possible to show (and it will be shown in the following chapter) that no viable alternative exists to the DEDRB evaluation methods for answering fundamental questions on the implementation and

effects of low-cost shelter interventions. 1/ Despite the general dearth of lending institutions or host country incentives to carry out evaluative investigations effectively, the current effort demonstrates that the record of such public sector interventions can be documented and assessed.

Although it is likely that the World Bank's role in future urban sector interventions will consist increasingly of macro-level programming, and decreasingly of administration of specific projects, the effectiveness even of such programming can be secured only by careful observation and analysis of project-level events, through coordinated efforts by Bank staff and national housing authority personnel. Program-level strategies are apt to miss at least some of their crucial targets unless a concomitant commitment is made to tracking the generation of project inputs through to their ultimate effects and impacts. Such tracking can help in the identification of areas where market-oriented interventions can supplement or supplant project-controlled components or activities.

1/ This is to be noted especially by those who continue to carp at the fact that these evaluations methods have been unable to start to produce results on health questions, and to claim as a result that the evaluation as a whole has been a failure. Such a conclusion is analogous to claiming that the automobile is not a useful invention because it cannot fly or operate under water. It is not enough to want health effects to be evaluated. It is also necessary to demonstrate that this can be done, estimate its costs, and decide to pay it.

Cost. The average cost of this evaluation program has been about US \$100,000 per country, per year. ^{1/} This total cost is substantial, but it is not out of line with similar comparative efforts at longitudinal evaluation undertaken elsewhere. ^{2/} Viewed proportionately, it comes to less than one tenth of one percent of the Bank's total urban lending (which has been US \$1.9 billion through FY81), and an insignificant fraction of total project costs of US \$4.3 billion. Evaluation thus represents a relatively small investment in research and development for a growing activity. Furthermore, the study's confirmation of many of the base hypotheses concerning the progressive development model permits future evaluation efforts to be much simpler and more operationally oriented, with possible savings of .5 to 50 percent in costs. It should be remembered that some questions, such as those about long-term impacts on household income, labor-force participation, and housing stock, not to mention health and education, can be evaluated only through longitudinal studies, and are therefore necessarily expensive.

Other Contributions. The evaluation program made numerous ancillary contributions which deserve acknowledgment. These include specific inputs to the projects under study, demonstration of the usefulness of monitoring and evaluation components within project

^{1/} This total excludes the cost of Bank staff time devoted to completing the research but includes all other costs of the program: the field costs, consultants' fees, supervision travel, and the Bank and IDRC contributions to the annual evaluation conferences/

^{2/} To the contrary, this has been a relatively low cost effort. (Ref. from the U. S. experience).

management information systems, development of empirical measures of project performance, methodological advances in the assessment of complex projects, and the progressive generation of new sets of issues for policy research.

In Conclusion. It is the combined judgment of research and operational staff that, on balance, the special evaluation program has been a success. Although the costs of the effort were significant, they were justified by the end results: a set of fundamental conclusions about a model of development that serves as a workable basis for the allocation of substantial resources in the urban sector in developing countries; and the development and refinement of a methodological approach to continue this process through future program iterations.

Related Lessons and Guidelines

The notion prevalent in the early 1970s was that the objectivity of evaluation findings could be safeguarded only by keeping the evaluation units apart from day-to-day project activities. The evaluation teams have found that, contrary to this notion, it is important that they be in close contact with project management in order to be of operational use, and to be on top of the actual course of project activities. ^{1/} In Senegal and Zambia, for instance, the original distancing of the evaluation units from project authorities

^{1/} The idea that evaluators should be kept close to project management -- and that management should contribute to evaluation design -- has also been advocated in recent years in the assessment of USAID's evaluation activities by Development Alternatives, Inc.; in the work of Joseph P. Wholey, reviewing other U.S. federal government evaluations; and by evaluation-oriented firms such as Abt Associates.

proved to be less productive than anticipated. As the teams developed a more integrated relationship with management, more of the evaluation findings served the pressing needs of the officers in control, including the review of selection (targetting) procedures; scheduling of meetings with project families on weekends or weekdays, as dictated by the work week; and providing crucial data on initial project impacts.

Placing of the FSDVM's Evaluation Unit in a direct reporting relationship to the Foundation's general manager, and the installation of the Research and Analysis Division in the Executive Staff of the Philippines' National Housing Authority, likewise forged strong and mutually beneficial links between evaluators and managers. However, it helped to utilize the option of occasionally subcontracting specialized elements of the research to individual consultants or local research institutions to obtain appropriate expertise. This approach, assisted by local research steering committees, permitted the evaluators to make the best use of local research capacities while maintaining objectivity. It was, in addition, necessary to ensure that the units were devoted strictly to evaluation activities, since in other development projects, including some financed by the Bank, using evaluators as the "eyes and ears" of project management led to their nearly complete absorption into routine operational roles. ^{1/}

These observations should not be taken to mean that evaluation staff believe that experience to date points to a rigidly fixed or

^{1/} There is also circumstantial evidence that evaluators' credibility as well-intentioned, unbiased, and objective observers has suffered somewhat under such arrangements.

standardized pattern for all relations between project managers and evaluation teams, though they do believe that functional and structural closeness between them facilitate optimal evaluation performance.

Additional important features have been coherent control by evaluation staff at the central lending institution, real supervision of field activities, and periodic conferences. Central direction of the evaluation efforts and their resources has been crucial. This structure of control has permitted adjustments in programs, as warranted by needs and opportunities within each program, as well as transfers of lessons from experience across countries. Flexibility in the management of budgets has allowed for desirable changes, as marginal productivities change, issues arise and subside, and it becomes essential to be able to move resources around within the evaluation program.

Our approach started out as one of multiple objectives, multiple instruments, and multiple types of studies, all framed within the format provided by quasi-experimental design. Born out of the Senegal project justification as impact measurement, we quickly added memory creation, calling it "giving projects memories." By the beginning of 1975, internal Bank discussions had already defined problem identification and resolution as equally important. Thus, evaluation came to be viewed not primarily as a tool for examining performance on behalf of project designers, but as a learning device to determine how and under what conditions sites and services and upgrading strategies can be most effectively used. The model that emerged was basically one of problem solving and record keeping that would lead eventually to good

impact evaluation. Some of the salient features of the model, as it has been applied, are enumerated below.

First, evaluation was conducted by a reasearch unit. This arrangement had the advantages of providing rigor and adequate resources, as well as a long-term horizon. It had initially the disadvantages of being too far away from operations, and of being led by the incentives that bear on research units to start at the "wrong end," that is, to come down from the full model rather than build up to it. (There was the added disadvantage that evaluation started out being an unpopular step-child in both parts of the house, being viewed as too time-consuming for operations, and insufficiently sophisticated for research).

A second salient feature was the approach chosen to finance the program. An early decision was made to go ahead with the program without resolving the mode of finance. It was decided initially only that financing would not come out of project funds, nor out of the research budget. This course led eventually to funding under a direct budgetary allocation, provided matching funds could be found from some other source; these were ultimately provided by the International Development Research Center of Canada. This independence of funding led to an independence of effort for the pilot period. It also provided some unique inputs through adoption in part of the IDRC method. This brought the advantages of extended dialogue with the field and of

networking among projects through conferences. ^{1/} A by-product was the fact that those involved learned to do collaborative research with developing country institutions much sooner and better than did some of their colleagues. Off-setting disadvantages included delays, some conflicts of objectives, and some administrative problems. On balance, this approach was excellent for a developmental phase. It would probably be less so in the long run.

A third feature concerns timing. It proved to have been the right thing to start planning the evaluation even before the projects started. Without the unforeseen delays in the Senegal project, the evaluation program would have been struggling to keep up. However, the delays in all projects ultimately had costly consequences of two kinds: part of the initial period was wasted in terms of gathering longitudinal information, and operations personnel did not welcome visible and substantial field activities that produced no results, even though it was clear why they could not be produced at that stage. Starting with a more operational approach and moving from bottom up would have been more appropriate to the circumstances on the ground, and better "politically" with respect to operations counterparts. This fact was recognized at the outset, but bureaucratic arrangements did not permit acting on it.

The distinguishing characteristics of the DEDRB evaluation approach are its comprehensive framework plus decisiveness in choosing issues to be worked on, the intensiveness of the effort in terms of

^{1/} For this component we are most indebted to the IDRC, both for introducing this activity into the program, and for financing the major share of its costs.

staff and talent, and the consistent willingness to keep the program flexible and evolving.

Close, periodic supervision has served to solve problems, make necessary adjustments, and keep individual field operations functioning. It has fostered joint learning about and appreciation of evaluation objectives and practices, as well as methodological interchange and improvement. The prime value of the conferences themselves is that they provided a forum for annual stock-taking, review, and modification of evaluation objectives and methods. This has been a great advantage, because, in spite of initial shortcomings of the approach and errors along the way, it has enabled evaluation management to keep adjusting the program to circumstances and interim findings, and thus to keep its workings at a high level of usefulness and efficiency.

PART IV

Chapter 9: The State of the Art in Evaluation Research, 1980

As of 1980, evaluation research has become a growth industry. During the 1970s, evaluation activities began to evolve into an independent professional discipline. A huge, two-volume, Handbook of Evaluation Research, edited by Marcia Guttentag and E. L. Struening, appeared in 1975. Three major new journals, Evaluation Review (formerly Evaluation Quarterly), Evaluation Studies Review Annual, and Evaluation and Program Planning joined the seminal Evaluation on library shelves, along with numerous "evaluation newsletters" produced by various government agencies. The Russell Sage Foundation devoted a large portion of its publishing portfolio to a Sage Library of Social Research, which focuses in part on studies in evaluation methods. An Evaluation Research Society was founded in Washington in 1976. And a number of graduate departments in university social science programs (such as Northwestern University's Psychology Department's Methodology and Evaluation Research Program) have begun to examine evaluation research as part of their curricula.

Even more significantly, the U.S. Federal Government has established Offices of Evaluation in major Cabinet departments, including State (in AID's Policy and Program Coordination Bureau), Labor and Health and Human Services. In fiscal year 1977 alone, some \$243 million was spent by government on evaluation of social programs,

one quarter of which went to support nearly 2200 full-time staff members with evaluation-related responsibilities.

Although the funds for such research may be reduced in future, it is clear that evaluation has become an institutionalized part of the federal program process, and on a substantial scale. And though most professional evaluators are still located in private consulting firms and universities, more and more practitioners are taking full-time bureaucratic positions.

Though growth is the salient aspect of evaluation activities in the 1970s, there is little agreement among professionals (who tend to be non-economist social scientists) as to the directions in which the field should grow. Whilst a lively dialogue continues among advocates of the four basic structures of evaluation research, these approaches remain largely points of departure. ^{1/} What is different today is the heightened awareness, after eight years' experience, of what evaluations can yield, given different resources of staff, methods, time, and program flexibility. Though the basic points of departure may be fairly entrenched by now, practitioners are better able to see them as alternatives in a range of options, now that the kinds of results each approach can produce are better understood.

Advocates of the rigorous experimental design, such as Robert F. Boruch, have also noted the usefulness of complementary, qualitative

^{1/} As discussed earlier, the four basic approaches are (1) controlled experiment design; (2) quasi-experimental and repeated time series designs; (3) correlational or co-variance techniques with some statistical controls; and (4) more process-oriented, quantitative strategies.

approaches to data collection and analysis, for program to data collection and analysis, for program evaluation. 1/ Similarly, forceful proponents of the "systems" approach have acknowledged the value of employing experimental or quasi-experimental designs, when institutionally feasible. 2/ Of course, there are still hardliners who reject anything but controlled experimentation, or even insist that such experimentation cannot be carried out effectively in the real world. 3/ Yet the fact is that the literature on evaluation has developed in such a way as to permit practitioners, observers, and users of research findings to foresee better than ever the probable results of different evaluation methods.

Current Directions in Evaluation Research

Three broad, overlapping trends can be discerned in current evaluation studies, all of which build on techniques already developed during the past decade. First, there is the strengthening of "modeling" paradigms to frame the research. This approach typically takes the form of specifying program implementation assumptions and procedures to chart expectations of project progress and weighs the relative importance of factors in the project environment that might dictate particular outcomes. A number of evaluative techniques have been employed under the rubric of "modeling", including fairly rigorous experimental methods

1/ Boruch and Reicken, 1975, p. 122.

2/ Mickelwait, Sweet, and Morss, 1979, pp. 167-179; Wholey, 1979, pp. 150-162.

3/ Campbell, 1979, pp. 80, 86-87; Pincus, Berryman and Glennan in Rand, 1980, Chapters 1 and 2.

(such as the Negative Income Tax Experiments), quasi-experimental procedures (including the Job Corps Program), and systematic process-oriented exercises. An example of each will be discussed below.

The second trend stresses clarification in advance of the uses and users of evaluation information. Much of the work in such efforts involves pinning down the linkages between program implementation and evaluation, and maximizing the utility of information presented to program implementors, participants, and other audiences. A variety of these techniques, including evaluability assessment, rapid feedback evaluation, and performance monitoring will be described and analyzed.

The third trend consists in refinement of measurements and indicators themselves. It embraces rapid (or rural) reconnaissance analyses and more sophisticated statistically-oriented techniques. Many of the technical analyses in the evaluation literature have been devoted to examining the strengths and limitations of particular experimental and measurement devices. A minor, noteworthy aspect of the third trend concentrates on identifying the procedures according to which re-analysis of evaluation, or "meta-evaluation," ought to be practiced.

The First Trend: Modeling and Implementation Paradigms in Non-Experimental Evaluation Studies

The continuing importance of the first trend, which stresses the tracking of project implementation, is due to the recognition that studies of goal attainment per se fail to capture the reasons why social or economic programs have the effects they do. Because American federal programming tends to dispense huge block grants to numerous sub-programs

in states and localities, evaluation efforts have gone to careful tracing of the allocations and how they have been spent. The increasing magnitude and expense of social programs, and legislative demands for accountability, have prompted evaluators to be more cautious and precise in their claims to measure outcomes and explain their causes.

Typical of the large-scale evaluation studies gauging the impacts of particular government interventions through social programs is the Rand Corporation's assessment of federally funded "change agent" programs in local-level educational systems. The Rand study was a four-year, two phase effort, begun in 1973. In the first phase, a representative sample of 293 schools in 18 states was studied for the effects of the innovative programs. The study relied on 1,735 personal interviews with staff at all levels in the school districts. During the second phase, 100 of the same schools were re-studied, and 1,343 interviews conducted, a year or two after the programs had been fully implemented and federal funding ceased. The evaluation attempted to document the nature and fate of educational "seed programs" as means of fostering educational improvement. The programs themselves varied widely among districts in terms of content, form, requirements for teacher training, and expense.

The Rand study (Berman and McLaughlin, 1978) found that the content of the educational innovations mattered significantly less than the manner of program implementation. Expensive efforts were no more likely to succeed than cheaper ones. The crucial factors leading to

program effectiveness appear to have been the active support of principals and local educational authorities, good working relationships among participating teachers, and the caliber of project directors leading the implementation effort. Particular strategies promoting mutual adoption of change agents led to the best results. These coordinated strategies included concrete, teacher-specific, extended training, with principals' participation and classroom assistance from project and district staff; teachers' observation of similar innovations conducted in other localities; development of teaching materials suited to local situations; and regular meetings on practical problems, with teachers participating in problem resolutions.

The recommendations that grew out of the Rand study focus on increasing attention to implementation procedures. The evaluators advise against emphasizing specific educational technologies, provision to districts of resources for their own sake, and even a targeted project focus, if these are not fitted to different districts' capacities to manage change. Rather, they favor flexible implementation assistance, allowing teachers, principals, and LEA's to meet needs as they arise in the pursuit of district educational goals.

Similar calls for understanding the local character of the socio-economic situations which development programs encounter have resulted from the national evaluation of the Follow Through Program, successor of the Head Start Program, and from the experience of Development Alternatives, Incorporated (DAI) with USAID's integrated rural development projects.

The DAI study merits discussion here as an example of implementation and modeling orientation in evaluation research. ^{1/} Based on numerous assignments in devising and coordinating inputs for rural development projects in less developed countries (though with little citation of specific efforts), the DAI study stresses that the organization of a development agency must be integrated and flexible. The DAI's research experience suggested that it is the nature of the development process that renders unpredictable the effects of even the most appropriate technological inputs, unless the organization of project implementation has been well worked out in advance. Organizational roles must be designed so that they incorporate incentives that make their execution feasible and worthwhile to development personnel.

The DAI report goes on to stress the need for both integration and coordination of development activities: "the measure of coordinated activity is ... the degree of information and resource sharing [required to guarantee the needed mix of goods and services], while the measure of integrated service delivery is the appropriateness (timing, quality, type, magnitude) of the mixture of opportunities received by the target population." ^{2/} The study cites instructive examples to show how neglect of these concerns has precluded realization of project goals.

^{1/} Integrated Rural Development: Making it Work?, Honadle, et. al., 1980. While not, strictly speaking, an evaluative research effort, Integrated Rural Development (IRD) is billed as a "preliminary review of the state of the art" of organizing and administrating IRD.

^{2/} Ibid., p. 31

One example is the programmed introduction of new smokeless stoves to improve respiration and eliminate eyestrain in Indian huts. One result was a much faster deterioration of the roof of the hut: termites clustering in the roof, no longer deterred by stove smoke, would make the roof's replacement necessary every six months, instead of every three years. What looked like appropriate technology caused the Indians to turn in their new stoves for old ones. In another instance, failure to plan for the time necessary to pay local level development officials eventuated in their spending 20 percent of their work time queuing for salary checks. The DAI study is overfull of examples of how failures of organization and procedural planning have thwarted the timing and outcomes of development activities.

The strength of non-experimental, implementation oriented studies such as those covered in the Rand and DAI reports is their insistence that social programs be carefully considered in the real context of their participants' lives, not in a vacuum created by theory. However, these studies suffer from their authors' reluctance to turn their findings and propositions into specific models for improving the kinds of social interventions they document. The Rand study identifies the crucial factors and circumstances necessary for installing a new educational program, but it stops short of framing these elements in testable propositions. The DAI studies are rife with propositions about what is and is not likely to work in development practice; but these propositions are never clearly adapted to on-the-ground situations practitioners must confront. DAI's discussion of

organizational problems concentrates almost wholly on failed aspects of projects, with little comment on what worked well enough throughout an entire project to justify replication of that effort. The Rand group and DAI might contend in response to this objection that a concrete model hinders the flexibility needed to deal with real problems as they arise, but this rejoinder would still leave unanswered the questions of what will work in a particular situation. The emphasis on process and on proposition formulation is no substitute for systematic, applied social theorizing in evaluation research.

Toward Modeling and Theorizing: More Rigorous Exercises in
Evaluation Studies

The last decade witnessed many serious attempts to apply the social experiment model to government-sponsored social research. ^{1/} Of these experiments, however, only two have dealt specifically with housing concerns: one local study examining the trend toward home ownership in the San Francisco area, and the better-known example of the Department of Housing and Urban Development's Housing Allowance Demand Experiments (HADE) to be discussed below. The majority of the social experiments were conducted in institutional or fairly easily controlled settings, such as prisons, hospitals, courtrooms, schools, etc. In such situations, randomization and the fitting of experimental variables to

^{1/} According to Robert B. Boruch (in Boruch, et. al., 1978, pp. 655-695), as many as 300 randomized field experiments have been conducted in the ten applied disciplines of criminal and civil justice; mental health, training and education; mass communications; information systems; research utilization; commerce, industry and public utilities; social welfare; health services and medical treatment; and fertility control.

control and treatment groups are more easily done than in developing country settings. In the latter kind of situation execution and even design of formal experimentation are extremely difficult due to political opposition to randomization of subjects into treatment and control groups, and problems with verifiability of selection procedures, and with attrition of samples.

1. Negative Income Tax Experiments The best-known series of social experiments conducted in the United States has been the Negative Income Tax Experiments (NITE). These experiments provided selected families with a guaranteed annual income as an alternative to current American welfare dispensations. The name of the program derives from the premise that when incomes drop below a specified level, income tax should become negative; that is, government should provide net revenue to citizens. Inaugurated at sites in New Jersey and Pennsylvania in 1968, and carried on for the next three to four years, the program became the subject of intensified debate in the mid-1970s. Analysis and re-analysis of findings from the original locations are still underway. Modified replications of the program were subsequently begun in Gary, Indiana; Seattle, Washington; Denver, Colorado; and other cities. Data analysis from those locations is continuing. The New Jersey study was expected to cost a total of \$8 million, \$3 million of which was allocated to participant payments and administrative costs, with the other \$5 million devoted to the research costs of program evaluation. Total investment in the NITEs comes to at least \$65 million, the largest sum yet provided

by government for a test of policy alternatives using rigorous social science methods.

In the New Jersey experiment, approximately 600 poor families of four members each, with working male household heads received income support payments bringing their annual income levels to between \$3000 and \$4000, or no higher than 150% of the established poverty level of that period. The funding was administered on four sites, according to eight distinct plans, differing in support levels and incentives for increasing family earnings. Another 600 families received no income support, but were given token remuneration for cooperating with quarterly interviews. The sites were selected at "test bores" in a few urban areas, rather than as a probability sample of some larger population of more direct policy interest, because the experimenters feared for the feasibility of a national sampling frame, wary of the inadvertent effects of welfare cases being included.

The research focus of the NITE's was to determine the disincentive effects of such income supports on male household heads' participation in the labor force. It was believed that poor families headed by working males would be most responsive to the disincentive effects of the experiments, and most likely to face peer criticism for receiving benefits. The experiments were not specifically designed to serve as a prototype for a negative income tax program, but rather as a research test case on income conditioned transfer programs. Participating families knew that they were the exceptional beneficiaries of a unique social welfare program, with guarantees of support for three

years. These factors would undoubtedly affect the generalizability of any results obtained.

The basic element of micro-economic theory being tested was that the two most important determinants of labor force participation were the guarantee of benefit level and the marginal tax rate. The use of the theoretical construct to policy makers and the specific measurements used will be discussed later.

The principal empirical result of the New Jersey NITE is that whilst income guarantees do not reduce the effective employment effort of poor working people, the final analyses do not reveal a substantial response to the two independent variables, the guarantee and the tax rate, around which the design centered. Individual responses to changes in income (and price) incentives turned out to be smaller and more refractory to analysis than the experimenters originally believed.

Valuable as the positive direction of this finding is, the inconclusive magnitude of the results hinders ready application of program procedures and objectives on a larger scale. Subsequent data generated from the later experiments, which have been based in part on the New Jersey program, may add to the interpretability of the findings.

The strengths of the design were its allaying of some significant threats to internal validity, through randomized assignment of subjects to treatment and control groups, and its testing of a facet of microeconomic theory with important policy implications. With regard to internal validity, care was taken to use census statistics on poverty areas and sample survey approaches in such a way as to minimize

selectivity biases. Instrumentation and statistical regression techniques were executed rigorously enough to reduce to some extent the possibility that the means of measuring impact variables would inappropriately affect the results obtained, and maturation as a factor in the experience of treatment and control groups does not seem to have produced adverse effects. As for the robustness of hypothesis testing, the measures used to check on labor supply response and wage income have been the subject of sharp debate (discussed below). It is generally agreed, however, that the analytical skills used to assess the data collected on the two groups were as acute as could be desired. ^{1/}

The exactitude of data analysis in the NITE has been overshadowed somewhat by basic challenges to other features of the design itself and by external events that may have seriously compromised the results obtained. For one thing, the design did not include measures to account for the fact that participants were receiving extraordinary support and considerable public attention, no complements to which were provided to the control groups. As a result, a "Hawthorne" or "guinea-pig" effect which could change drastically if the NITE became the basis of a national income-support policy. The New Jersey experimental group was singled out for attention by network television, a factor that may well have motivated "good behavior." Also, randomization procedures were executed on the basis of individual families, rather than neighborhoods, maximizing the contrast between families receiving benefits living alongside others who received none. Attrition rates were in

^{1/} Rossi and Lyall, 1979, pp. 425-426.

fact high for the control group, reaching 25.3 percent over the three-year period, as compared with only 6.5 percent in the most remunerated experimental group.

A second objection to the design relates to the provision of support with a built-in three-year time limit, which may have prompted participants to keep working in their present jobs (unless better ones were found), in order to be on a more favorable footing when confronting the more uncertain environment following the experiment. Entitlement to income support on a different, more discontinuous, or bounded schedule might well have produced different responses to job opportunities. Such criticisms challenge the reliability of differences in magnitudes of measures applied to experimental and control groups. ^{1/}

Turning to the practice of using "test-bores" as a site-selection procedure, critics aptly complain that insufficient attention was given to defining the universe which the four sites were purported to represent. ^{2/} The absence of measures for reckoning the differences among sites, they point out, hinders generalizability of the experiments' findings to sites elsewhere in New Jersey and Pennsylvania, let alone the rest of the nation. Because the eligible population was defined as those male-headed, intact, working families earning no more than 1.5 times the prevailing poverty ceiling, a truncated sample resulted, especially for whites, in which family size was greater than average for the poor population in general, with fewer working wives,

^{1/} Campbell, 1979, p. 81.

^{2/} Rossi and Lyall, 1979, p. 416.

lower levels of home ownership than anticipated, etc. Such truncation not only affects the applicability of the results in the policy arena on which the experiment focused, it skews the minority- or ethnicity-specific effects that a more disaggregated design might have encompassed.

A final, admittedly unforeseeable, factor further limits the utility of the New Jersey NITE results. When the experiment began, that state did not offer a welfare program to intact families headed by male workers. However, within a few months of NITE initiation, this very provision was incorporated into a state Aid to Families with Dependent Children program. Rossi and Lyall contend that:

"Such changes...did affect the nature of the [NITE] experimental treatments, subjecting them to competition (CK?) from welfare policy that changed the experiment to one that measured the effects of NIT[E] on work response when added to generous AFDC-UP plans. As such, it was likely to underestimate the impact of NIT[E] considered alone and also when used as a supplement to less generous AFDC-UP plans...[O]verlaying the experimental treatments with a competing AFDC-UP program made it much more difficult to determine in the final analyses whether the (experimental) treatment was the nominal guarantee and the tax rates, the difference between these and the competing AFDC rates, or some even more complex combination of rates with "kinks" at points where there occurred incentives to switch from one program to another." ^{1/}

In short, if the external validity of the NITE faces severe limitations, critics question as well the extent of the experiments' internal validity. In light of changes affecting the inner structure of the income support program itself.

Rossi and Lyall add another cautionary note regarding NITE implementation. They aptly maintain that, if "treatment" in a social

^{1/} Ibid., p. 418.

experiment includes all that is "done" to an experimental group that is not "done" to the control group, then the groups' experiences of the administrative program itself have to be considered part of the experiment's design. Whilst Rossi and Lyall acknowledge that the experiments tried to minimize their own contacts with participant families, entrusting the consistency of administrative contacts with families to field office personnel, television interest in particular families and other external investigations of the NITE may well have led to differential responses on the part of administrators as well as participant families. Such potential administrative effects were not foreseen, and no provision was made in the design to account for them. Yet, possibly comparable effects have been known to be carried from one NITE to another. Rossi and Lyall note that the administrative experience of the Scranton experiment was affected by the fact that the Pennsylvania staff spent nearly a year "learning by doing" in New Jersey. Because the NITE will be analyzed by scholars, without explicit stipulations made for the range of different administrative procedures, results might well be assessed without a full recognition of the behavioral variations inhering in "standardized" administrative packages. Accounting for each of the possible threats to internal, external, construct, and statistical conclusion validity certainly helps to minimize differences in treatments, but even a quantitatively-oriented practitioner like Donald Campbell admits that "many of the methodological problems of social project implementation and impact measurement have to do with the social psychology of the interaction

between citizens and projects, or between citizens and modes of experimental implementation...or between citizens and the special measurement procedures introduced as part of the evaluation." ^{1/}

Despite the real problems involved in the NITE design and implementation, Rossi and Lyall readily admit that the program raised substantially the level of policy debate on experimental techniques and their relation to federal programming. They believe that the relatively high cost of such experiments is justified when the importance of the policy subject and the inadequacy of existing survey data combine to merit the application of the complex procedures involved. Several leading social policy makers concur with Rossi and Lyall that the Negative Income Tax Experiments underscored the potential, and perhaps the need, for testing alternative administrative delivery systems for social programs.

2. Experimental Housing Allowance Program. Another rigorous evaluation exercise in American federal programming has been carried out in the U.S. housing sector. Under the umbrella title, Experimental Housing Allowance Program (EHAP), this Department of Housing and Urban Development program has three principal components: the Housing Allowance Demand Experiment (HADE), the Housing Allowance Supply Experiment, and the Administrative Agency Experiment (AAE). Aspects of the HADE come close to the work reported on in the DEDRB evaluation, but results from the AAE interestingly reflect the discrepancies between program planning and implementation as discovered by evaluation

^{1/} Campbell, 1979, p. 68.

efforts. Each will be covered in turn, the first because it represents the major rigorous American investigations of housing demand, and the second because it brings into focus the contributions quantitative and qualitative approaches make to evaluation research.

The HADE program, carried out among 1700 households in two urban sites in the period 1973-1977, examined the housing demand behavior of several different low-income experimental groups and a low-income control group. Specific housing allowances were offered to the experimental groups as a way of measuring price and income elasticities of housing demand. Households were offered a variety of arrangements, depending on their random assignment to the groups: 1) a percentage rebate on rent payments, with no specific housing requirements; 2) an income-based housing allowance conditional upon meeting a "minimum rent" payment requirement; 3) an unconstrained housing allowance payment, similar to the general income subsidy used for the NITE; 4) an allowance payment only if the household's dwelling unit meets minimal housing standards; or 5) participation in a control group.

Analysis of HADE program effects is being carried out for HUD by the consulting research firm of Abt Associates, which is still gradually reporting results. Two associates at Abt, Joseph Friedman and Daniel H. Weinberg, have compared effects of granting allowances to the first "percentage of rent" experimental group (with no housing stipulations) with a control group receiving no stipend. Two housing expenditure functions were used to calculate elasticities, a log linear and a linear form (derived from a Stone-Geary utility function).

Friedman and Weinberg found the price elasticity of housing demand for low-income renters to be $-.22$; that is, a ten percent price reduction prompts a 2.2 percent increase in housing demanded. Permanent income elasticity is estimated to be $.36$ (or, a 10 percent increase in income yields a 3.6 percent increase in housing purchase). ^{1/} These elasticities should not be thought to be constant over the housing price and income range. Rather, as suggested also by the DEDRB evaluation on affordability of low-income housing, the elasticities are likely to increase in absolute value as income increases. ^{2/}

Friedman and Weinberg have also gone on to investigate HADE program effects on the other three experimental groups, again measured against the control group. Because the entire sample of households was selected on the basis of households' pre-program housing consumption, a variable endogenous to the experiment, the two Abt analysts correct their response estimates for selection bias. ^{3/} It turns out that pre-program status had a highly significant effect on the way households responded to the housing allowances. Those households that met housing requirements at enrollment, and were thus automatically eligible for allowance payments, did not use that allowance to increase housing consumption on expenditures. Evidently, these households, not experiencing major housing deficiencies, used the stipend for other

^{1/} Friedman and Weinberg, cited in Stromsdorfer and Farkas, 1980, p. 443.

^{2/} Jimenez and Keare (Article).

^{3/} Adjustment for the resulting bias led to lower treatment effects for program sites (in Phoenix and Pittsburgh) across all treatment variations.

needs. However, households that fulfilled minimum rent or standards requirements only after enrollment, and apparently did previously face serious housing deficiencies, made large housing expenditures as a result of participation in HADE. These expenditures did not exhaust the allowances awarded, and residual resources were evidently used to satisfy needs other than housing. Another major finding was that the unconstrained allowance accounted for roughly the same increase in housing services and expenditures as did the minimum standards and rent requirements. Variations between program sites in Phoenix and Pittsburgh, unfortunately, could not be fully explained by the analysis, though differences are apparently due in part to dissimilarities in the two cities' housing vacancy and mobility rates. 1/

Other HADE-Abt investigators have employed different methods to interpret the housing allowance data. Jerry A. Hausman and David Wise began from the premise that the housing allowance offered created a discontinuity in the budget constraint felt by the experimental groups. They thus posited a Cobb-Douglas utility function with the share of income going to rent expressed as a function of household characteristics, including income. With this model, Hausman and Wise estimated housing demand elasticities to be $-.16$ for price and $.6$ for income, calculations similar to those derived by Friedman and Weinberg. 2/

Evaluators Ernst W. Stromsdorfer and George Farkas note that constrained housing subsidy programs are not the only means to aid poor

1/ Friedman and Weinberg, 1980, p. 446; Stromsdorfer and Farkas, 1980, pp. 443-4.

2/ Hausman and Wise, 1980, pp. 478-479.

families in increasing their housing stock and raising its quality. Unconstrained income maintenance programs, such as the negative income tax, are another type of vehicle for promoting social ends such as housing enhancement. As evidence of this, Stromsdorfer and Farkas cite statistics from the work of James C. Ohls and Richard L. Kaluzny on housing demand in the Denver/Seattle and Gary Income Maintenance Experiments, respectively. For Gary, where the experimental sample was divided into renters and owners, Kaluzny found that the estimated rent elasticity was approximately .30; that is, a 10 percent increase in income support yielded a 3 percent increase in rental housing purchased. Ohls found comparable results in Seattle and Denver to be .42 and .32 respectively, thus making the NITE results for housing generally consistent with those of HADE.

The similar elasticity estimates do not mean that the absolute expenditures induced by general income support and specific housing allowance programs will be the same. To the contrary, families can and do spend less on housing from general income subsidies than they spend when given specific housing support. For example, EHAP participants increased their net rental expenditures by 19 percent, however only a 5 percent increase resulted in the Gary NITE. As Stromsdorfer and Farkas observe, however, in-kind subsidies have their value, but are a less flexible resource to beneficiaries than an unconstrained income allowance of equivalent value. The Abt researchers believe these findings argue for the use of specific housing requirements only in cases where policy makers have specific objectives in mind that clearly

call for such measures. Housing stipulations evidently compel subsidy recipients to seek housing that satisfies the requirements, while otherwise providing for only the normal level of housing quality per dollar spent. Although poor families may conceivably improve their circumstances within such program constraints, their circumstances theoretically could not be improved by more, and could probably be improved by less than what could be achieved through an unconstrained income support program. It may still be that the optimal housing subsidy approach for low-income Americans is a program combining a number of options, including supply-oriented features, such as subsidized public housing. The DEDRB evaluation has similarly concluded that national welfare authorities must weigh alternative methods of meeting housing and other basic needs according to the relative efficiency and effectiveness of means to enhance housing, health, nutrition, etc., directly, vis-a-vis untied support programs.

Like the NITE, the HADE represents a major American investment in testing alternative social assistance possibilities. Friedman and Weinberg report that Congress has spent over \$200 million on research and development of low-income housing programs. More important still is the fact that the HADE built upon housing data collected in the NITEs: a fairly controlled comparison has thus been developed between the effects of subsidies in kind and unconstrained cash payments as methods of upgrading urban housing stock. By testing various microcomponents of housing allowance schemes, and assessing the relative micro-expenditures of income maintenance program beneficiaries among different household

needs, social experimentation has produced a sound basis for constructing a macro-economic perspective on the satisfaction through public programs of the basic needs of low-income Americans. Though conclusive results on alternative subsidy interventions have yet to be generated, research has provided policy makers with increasingly sophisticated and generalizable data with which to work. The limitations to date on policy innovation resulting from the experiments derive as much from the lack of political consensus about evaluation findings as from any particular methodological weakness of the studies themselves.

3. Administrative Agency Experiment: The apparently contrasting results of qualitative and quantitative approaches to program analysis can be illustrated by reviewing an experimental housing allowance program, known as the Administrative Agency Experiment (AAE), which was coordinated with HADE. The experiment, devised and run for HUD by Abt Associates, attempted to identify optimal program management strategies for direct cash assistance approaches like those of HADE. Eight public agencies across the country were to create their own housing allowance programs, each with authorization to help seek upgraded housing services for up to 900 families over a period of two years. These independent agencies were then subjected to Abt Associates' quantitative and qualitative assessments of progress, to ascertain whether cash dispensations were actually being used for enhancement of housing services consumption. However, the administrative agency staff at the case study site selected for in depth study was apparently not fully

apprised of the fact that full program funding (for the beneficiaries' housing allowances, as well as staff salaries and administrative expenses) was contingent upon maximizing the number of families placed or maintained in upgraded housing units. It turned out that the resulting disjunction between quantitative analysis based on in-course obstacles and outcome measures, and qualitative analysis based on the process of handling clients, derived from a misunderstanding of how agency unit costs were actually being allocated to processing and placing clients--a measure inadequately grasped by both quantitative and qualitative analyses. 1/

Additional descriptive details are necessary to flesh out the significance of the case. This administrative agency's jurisdiction covered three largely non-contiguous areas, two rural and one urban. Each contained a branch office with staff to interview candidate families and inspect, or arrange for inspection of, the original and improved housing. Few direct social support services were to be offered, however: after a brief explanation about how to locate better housing, each enrolled family was responsible for finding the unit desired. 2/ The parent public (or "contracting") agency was to monitor progress largely by the number of households receiving payments at a particular time, with benefits available only to households that had met

1/ Trend, 1979, pp. 68, 79-82.

2/ Families could make improvements in their original dwellings and qualify for program benefits, provided inspections were passed.

the standard requirements within 90 days after enrollment. ^{1/} Other quantitative measures besides fiscal accounting were also used to keep track of expenditures. Qualitative analyses of administrative agency performance were to be carried out by an Abt-appointed observer who assessed the nature and effectiveness of the "chain" of services, from initial interviews to family placement in upgraded locations.

Whilst rural enrollees seemed to meet program requirements easily from the outset, urban families soon began dropping out of the program in large numbers, without actually qualifying for subsidies. The contracting agency thus pressed the administrative agency to meet daily processing quotas for handling enrollees and to minimize contact between staff and families. The support services-oriented staff complained of being overworked by "paper pushers" and of not being given sufficient time to ascertain that recipients had not merely found or stayed in "junk housing." The conflict between contracting and administrative agencies was further aggravated by the former's aim of making the enrollee population racially representative of the site area, whereas the latter claimed that such a position led to underenrollment of black families, who were attracted to apply for the program in large numbers. Administrative agency staff members repeatedly argued that they preferred meeting the needs of 500 families well to pushing through 900 families who may or may not in the end have received the amounts or kinds of benefits to which they were entitled.

^{1/} As it happened, the staff at the case study site was social services oriented and did not agree with the approach of monitoring progress in terms of the numbers of families enrolled and placed.

Contract agency personnel on the other hand insisted on meeting numerical and demographic objectives.

At the end of the program's implementation, 900 families had, in fact, received housing allowance payments, but not before many administrative agency personnel had resigned in protest over the "assembly line" approach imposed upon them. Initial quantitative analyses of the eight cross-country sites did indicate, though, that the case study site had been an efficiently run program, with a representative minority population, and that beneficiaries had upgraded their housing to the second-highest degree within the AAE as a whole. Analysts of the qualitative performance, however, insisted that the conflict between the agencies, the managerial "incompetence", and the concentration on placing families in housing at minimal cost, had severely compromised the commendability of the site's performance as a model of a housing administrative program. Because the qualitative and quantitative assessments were so at odds, M. G. Trend, another Abt evaluator from headquarters was brought in to help reconcile the differences in findings from the two approaches.

Trend set out to examine the conflicting claims of the contracting agency and the housing program (administrative agency) staff that had given rise to the dispute. Although the observers' records indicated that complaints about pushing caseloads had begun early in the project, cross-program comparisons revealed that the workload was not higher on balance at the case study site than elsewhere. In fact, non-complaining rural staff had processed almost as many households, on a

per-person basis, as the complaining urban staff even within the three-area site.

Trend went on to examine other differences that might emerge from an urban-rural comparison within the site which had previously been treated as an undivided whole. Because it was known at Abt headquarters that funding for housing allowances, and staff and office expenses came from the same "pot," Trend hypothesized that the pooled-funding procedure may have inadvertently generated some competition for funds. It was at this point that he discovered that administrative agency staff had not known that serving any less than the full complement of 900 families denied staff part of their own obtainable program funds. It further turned out that recipient household size specifically affected the program cash flow to beneficiaries and to staff workers. Because the rural areas contained smaller households with larger incomes (and more whites) than the urban site, rural families on average received lower payments than urban families--but more funds were made available at the same time for the rural program's administrative and staff needs. In short, money was tighter at the urban site than at the rural site, or elsewhere. Ironically, the high turnover of black enrollees kept more funds within the urban program than if more black families had actually received subsidies. Because black families tended to be larger than (urban or rural) white families, their larger numbers would have diverted a larger proportion of funds from the urban program administration to the housing subsidies themselves. The resignations of

staff members also kept more funds in the urban program than would otherwise have been available.

Finally it was discovered that the contracting agency's penchant for checking the numbers of households eligible for receiving funds included only one thorough termination of those households who had exceeded the 90-day limit in finding new housing. Technically, then, the quantitatively oriented monitors "could not have known how many people were looking for housing at any one time or how they were faring." ^{1/} Meeting payment objectives was no substitute for knowing the ongoing success rate of participant families trying to find new housing. Moreover, because the program efficiency rate was based on cost per enrollee, not on the cost of enrollees less drop-outs, the site in question, with its many drop-outs, was able to appear more efficiently run than it actually was. The failure to calculate the wastage factor might have been detected sooner if Abt headquarters had heeded the staff complaints that the total volume of applicants was not being served as effectively or efficiently as possible. It also turned out, ultimately, that the housing quality measure used by the program designers was itself unsatisfactory.

Trend's investigation showed that the conflict of evidence at the AAE site was not so much a result of disparate or opposed data bases, as it was a contrast of accounts based on different perspectives on different kinds of data. While his re-analysis does not fully answer the questions remaining about the AAE site--the real extent of the

^{1/} Trend, 1979, p. 80.

administrative program's efficiency; the tradeoffs incurred in stressing recipients' placement in housing, rather than the provision of other housing support services; the scope of eligible clients reached; etc. -- Trend's work does demonstrate the usefulness of triangulation as a key aid in large-scale social programming. He does attempt to bring a variety of data and means to bear upon the same problem. If the result does not constitute a full synthesis of qualitative and quantitative methods, which it does not come close to doing as he himself notes, Trend at least reveals the strengths and weaknesses of the two approaches in clarifying the AAE findings, and leaves it to the reader to decide what is to be learned from the experience.

4. Job Corps Evaluation: The ongoing Job Corps Evaluation provides another fairly rigorous example of the measurement of social program effects. The program, a cornerstone of the Johnson Administration's War on Poverty, provides job-skills training, fundamental education, work experience and placement, and health care to severely disadvantaged youths in residential settings removed from the "debilitating" environments plagued by the severe unemployment problems the program attempts to alleviate. It was incorporated as Title IV of the Comprehensive Employment and Training Act (CETA), passed in 1973. Some 45,000 enrollees were provided for by the first federal budget, with the yearly cost per per Corps member amounting to slightly more than \$10,000. ^{1/}

^{1/} Kershaw and Maller, 1980, p. 3-4,

The model of program effects constructed on the Job Corps Evaluation, by the firm, Mathematica Policy Research (MPR), posited that the cited program features would both improve future earnings potential and generate attitudinal preferences for labor force participation and education. Measurement of program effects centers on four categories: 1) employment and earnings; 2) investments in human capital; 3) dependence on welfare and other public transfer programs; and 4) anti-social behavior. The evaluation hypothesizes that, relative to non-participating youths of similar socio-economic background, the Corps experience: 1) helps create more employment and higher earnings; 2) promotes more returns to school or other training programs, more qualified applicants for military service, etc.; 3) reduces receipts of cash and in-kind transfer payments; and 4) results in lesser incidence of criminal activity or dependency on drugs and alcohol. Table 9.1 presents a summary of the evaluation's proposition testing.

Because the MPR evaluation was initiated after the Job Corps program had begun, the firm was obliged to forego using a fully controlled experimental design using random assignment of individuals to program and non-program groups. Instead, a quasi-experimental comparison group was matched to Job Corps enrollees, based on similar socio-economic backgrounds, including school dropout status and registration at local employment security offices. As a second feature of the comparison group, selection effect probabilities were reduced for areas already saturated with Corps members, or for those areas close to the Corps centers themselves. The assumption was that those comparators

Table 9.1: SUMMARY OF HYPOTHESES FOR JOB CORPS PROGRAM
IMPACTS ON PARTICIPANTS

Relative to if they had not gone into the program, participants will:

a. Employment and Earnings

1. Have more employment
2. Have more stable employment
3. Have higher earnings
4. Have higher wage rates

b. Investments in Human Capital

1. Be more likely to return to school or continue their education in other ways
2. Be more likely to be in higher levels of education
3. Be more likely to enroll in training programs
4. Be more geographically mobile
5. Be more likely to qualify for military service
6. Be more healthy

c. Dependence on Welfare and Other Public Transfers

1. Have reduced receipt^s of cash transfer payments
2. Have reduced receipt^s of in-kind transfer payments

d. Antisocial Behavior

1. Be less likely to engage in criminal activities
2. Be less likely to use drugs or alcohol

X

Source: Karshaw and Mallar 1980:7)

living in such "saturated" areas, or close to Corps centers which had disseminated much program information, were more likely to have "self-selected" themselves out of participating. A baseline questionnaire was also used to complete the matching and to model any differences that remained in the comparison group after matching. This screening procedure helped establish actual eligibility for Corps membership among the comparison group members. Finally, statistical techniques devised for the analysis controlled for both observed and unobserved differences between Corps members and comparators, including unseen potential differences such as job motivation. Few significant differences were found between the sample groups when they were examined for race/ethnicity; age; family type; marital history and other demographic factors; education and training; parents' socio-economic background; previous and current income; and incidence of antisocial behavior. 1/

Sample selection and maintenance also demanded considerable attention: 10,000 interviews had to be conducted at 52 Job Corps centers and 15 comparator sites all over the country in a period of 24 months. Initial Job Corps interviews were not especially problematic, though neither were they negligible in cost: 6600 preliminary interviews were conducted in person, with subsequent waves making use of direct telephone, and mail communications. 2/ By clustering enrollees into groups by home geographic region as separate probability sampling units, and deriving "the target sample size from estimates of the analytic

1/ Ibid., pp. 9-10.

2/ Such mixing of information gathering techniques is in itself a questionable procedure.

precision required and consideration of sample clustering and sample attrition over time, the sampling rate was determined, ...represent[ing] a typical two-stage area probability sampling solution to the problem of minimizing survey variance subject to a budget constraint." 1/ To maintain the sample, conscientious efforts were made to stay in touch with relatives of the young respondents who would know their whereabouts; to keep interviews short, relevant, and unthreatening, whilst at the same time emphasizing the usefulness of the respondents' honest answers; and to provide monetary incentives (\$5 payments per interview) to keep respondents in touch for re-interviewing. 2/

Preliminary results of the evaluation have been obtained for Job Corps program effects at a number of intervals after members had completed their training. On average, these results derive from interviews conducted seven and then 18 months after program completion. It appears that the Job Corps Program is generally successful, for both men and women, in: " (1) increasing employment and earnings; (2) improving future labor-market opportunities through work

1/ Ibid., p. 11.

2/ The interviewers themselves are largely from minority groups, a third of them being under 25. They were trained to elicit as much information as possible and to reassure respondents of the interview's confidentiality. Interviews were conducted at locations minimizing the chance that answers might be overheard by family members whose presence might have inhibited frankness. "The impacts of monetary incentives on the success of the search effort (for reinterviewing) increased over time ... For every hundred sample members paid .. we received ten more responses to an advance mailing than for a like number of (unpaid) sample members, which shows some clear advantages of respondent payments in terms of search efficiency and completion rates of mail surveys." (Ibid., p. 13).

experience, education, better health,... and military service; (3) reducing dependence on welfare and other public transfers; and (4) reducing criminality. Underlying the moderate magnitude of these findings are extremely large and statistically significant estimates for program completers and much smaller, insignificant estimates for Corps members who drop out of the program before finishing." 1/ While unobserved differences between program completers and non-completers may account for part of the statistically significant difference in these effects, the evaluation claims that these differences are in turn at least partially attributable to full course participation in the Corps program. 2/

The timing of the desired program outcomes is also particularly noteworthy, especially those related to employment and earnings. Corps members evidently undergo a period of readjustment immediately following program participation, for they initially experience greater levels of unemployment than non-participants. Thereafter, however, employment status increases to significantly higher levels. As MPR notes, the "delayed reaction" to participation may have been misinterpreted if the evaluation had confined itself to post-program placement measures in the months immediately following completion. Different results may be obtained depending on when and over what length of time program analysis is performed. 3/ In

1/ Ibid., p. 19.

2/ Ibid., p. 21.

3/ It will be recalled (see pg.) that the DEDRB has recorded, as an illustration, a similar finding with respect to rental housing. Many more examples could be cited.

particular, cost benefit analysis might undervalue positive program effects unless measurements extend beyond initial readings such as of initial job placement. 1/

MPR is admirably candid about the limitations of the Job Corps study, limitations that apply to most, if not all, evaluative efforts. Evaluations are commissioned to capture program effects at particular times, but these readings of effects may not encompass the full range of impacts. Policy makers must thus be alerted to the fact that more effects may be engendered by social programs than are measured, or measurable, by evaluations conducted during or in the immediate wake of program activities. MPR also notes that the social and economic context in which a program is carried out is dynamic, whereas rigorous experimental and quasi-experimental designs are largely static constructions which need to be re-devised if significant program changes occur. Hence, evaluators need to be watchful of both treatment and comparison groups, as these are both likely to undergo different rates of social and individual change during the experimental period, and beyond.2/

As carefully executed as the MPR Job Corps evaluation was, it has not identified all the intervening variables that linked program

1/ Education evaluator Lois-Ellen Datta has also maintained that focusing on the longer term can yield more positive program results than evaluations gauging only short-term impacts. She notes, as well, that training and other work programs which do not assess the value of work performed during training (both in terms of actual products and gains in human capital, such as skill levels) miss at least part of the real output of such programs.

2/ Ibid., pp. 24-27.

inputs to successful program outcomes. In this respect, it recalls some of the liabilities of the Rand and DAI studies. MPR does not weight individually the range of factors (training, health care, job assistance, etc.) that are said to be responsible for the socially and individually beneficial program effects. Admittedly, the study reviewed is oriented toward the economic impacts of the Job Corps program. But if program replicability or expansion are likely desiderata, the evaluation ought to cover the relevant factors yielding program success. What kinds of training (content and methods) led to the earnings and employment results in the short run and in the longer term? If future programs are not allocated the same high level of resources as the Job Corps, what factors are crucial to achieving similar outcomes through a program of reduced scope? Must Corps members be isolated in reserve centers to attain the good results? Would other program methods have achieved the same results as those used? Would other evaluation methods have determined the outcomes to the same degree of precision?

These are the sorts of questions the MPR study has yet to address with its "pure" economic analysis. Given that the evaluation is still in progress, it may yet attempt to do so. It is important the effort be made in that the multi-faceted Job Corps program would benefit from "triangulation" of economic, sociological, and social psychological investigations of the full effects of program processes.

Interim Assessment of the DEDRB Evaluation vis-a-vis Trend I

The foregoing summaries of major American social program evaluations constitute one view of the state of the art in the trend of

evaluation research toward modeling and theorizing. In our view, the DEDRB evaluation compares favorably with these efforts. This assessment is based on several factors: the clarity of the urban shelter program research design, which specified the independent, intervening, and dependent variables in the project environment; the quality of project progress tracking over time, which revealed the interaction of these variables; and most importantly, the theorizing that resulted from the tracing of links between project inputs, outputs, and effects.

As the above discussion has stressed, few American evaluation efforts have included the modeling of formal hypotheses postulating how and why particular results have occurred. The examples cited tend toward varying degrees of acceptable hypothesis testing, but with one exception do little to replenish theory (of instituting educational innovations, integrating rural development, or measuring disincentive effects of income support payments) by offering little or no new modeling of results for future studies. Put another way, the DEDRB evaluation makes a positive contribution to the body of knowledge that supports policy formation in a specific social sector. Most comparable evaluation research reveals a tendency to shy away from using particular study results to formulate new strategies in the relevant sectors.

The exception is the contribution of Abt Associates to the analysis of housing demand based on HADE and AAE program effects. Like the DEDRB evaluation, those efforts have posited a formal design for assessing changes in variables through a project's course. They then chart the development of what happened in the course of project imple

mentation and model the results in a testable fashion. Because those contributions are based on specific issues in sectoral research, linking behavior to supply and demand manipulations within the framework of a discipline, they both fit within the discipline's ongoing growth, and enlarge its scope by specifying new directions for future studies. For instance, given the rudimentary state of the art of hedonic analysis of housing quality and value, any contribution in this area furthers the work of detailing how particular constellations of housing characteristics can be translated into demand functions. The fact that the theoretical results of the DEDRB evaluation complement those of the HADE and NITE in design and rigor, marks them as pioneer efforts in specifying elasticities in housing demand across income ranges. The same can be said of the DEDRB contributions to theory about the affordability of housing to low-income populations.

The Second Trend: Clarifying Evaluation Uses and Specifying Its Users

Practitioners of the first trend in evaluation, discussed above, have concentrated on maintaining the autonomy of evaluation methods and findings from program constraints. These researchers have not limited themselves to serving the needs of program managers alone, but have been concerned as well with contributing to general social science research theory, methodology and empirical findings. ^{1/} A second trend in evaluation research focuses strictly on efforts to make evaluation procedures and results useful to project implementers, program managers, and higher level policy makers, all clearly identified

^{1/} A central concern, as well of the DEDRB evaluators.

in advance. The premise of this approach is that policy-oriented research that is unused or ignored by program controllers is largely or wholly wasted, that evaluation information useful only to research colleagues is not acceptable. The majority of evaluative techniques developed to serve specific program audiences stress the rapid utility of information collected, but in some cases have failed to answer direct questions posed by program users. In other, smaller-scale, more issues-oriented studies, evaluations have satisfied their intended audiences. Cases of both kinds will be discussed.

User Orientation.

First, we will look again at the Negative Income Tax Experiments with an eye to their consideration of specific consumers of research information. The experiments were clearly of interest to a number of important "users": Congress, the federal administering agency, researchers, and the general public. Congress hoped that these income support experiments would shed light on the relations between income guarantees and labor force participation, and on the costs of mounting a national program if the work disincentive factor proved to be slight or negligible. Records of legislative debate also show that Congressmen were interested in knowing the generalizability of research results to female-headed households with different employment circumstances, as well as the effect of income supports on intact families headed by male breadwinners. Policy weights of different NITE features were devised without substantial reference to the specific concerns either to the program's administering Office of Equal

Opportunity (OEO) or to Congress. While OEO and its parent agency, the Department of Health, Education, and Welfare (HEW) kept close tabs on the general import of the NITE research, they apparently chose not to interfere with the technical aspects of the experiments' design. It appears too that the exPerimenters did not consult often with other evaluation researchers on the adequacy of their chosen indicators of such measures as earnings, income, wage-rates, and hours worked.

Peter Rossi and Katherine Lyall clearly view the New Jersey NITE as inadequately designed to satisfy its most vociferous audiences, above all Congress:

"Most [legislative] participants in the F[amily] A[ssistance] P[rogram] debate regarded the [preliminary and final research] findings as "inconclusive," biased by the interests of the researchers, or failing to speak of the active issues of political concern about the FAP. Virtually no interest was shown in designing or fine-tuning a national program using the tax and guarantee rates examined by the NIT[E]." 1/

Though these commentators do not claim that the NITE promised national cost estimates for low-income family support programs, they do maintain that this was OEO's primary interest in the research and that the limitations of the sample restricted the technically sophisticated advice which might have been brought to bear on policy debate in the national political arena. 2/ Although Congress did go on to spend millions of dollars on additional research-oriented income maintenance programs, it stopped short of approving any sort of program or policy that would have authorized family supports on a national scale. Rossi

1/ Rossi and Lyall, 1979, p. 427.

2/ Ibid., p. 415.

and Lyall, Donald Campbell, and other critics concur that Congressional examination of the NITE results did little to bolster the reputation of rigorous social science research as a tool to inform the lawmaking process.

However, the formal experimentation and other exacting research procedures employed have impressed the most directly concerned federal program staff concerned with commissioning and disseminating the findings of social science research. Rossi and Lyall admit that, though the NITE research may not have raised the level of Congressional debate,

"...the results were extensively discussed within HEW and that a number of HEW officials were willing to revise substantially this belief (in large disincentive effects) in the face of experimental results... The academic community has been fascinated by the technical design contributions of the experiment and the demonstration that experimentation is an administratively feasible research technique, but has been correspondingly skeptical of the results in light of the kinds of design flaws discussed." ^{1/}

In short, the NITE research impressed some of its audiences and disappointed others.

It is often difficult to determine the degree to which, or the specific ways in which evaluation results have affected policy decisions, especially decisions involving the disposition of tens or hundreds of millions of dollars in large scale social programs. HADE, for example, has been tested on at least as sound a theoretical and methodological basis as the NITEs. But it is far from clear to what extent site-specific analysis problems, budget constraints, program and administrative shake-ups, or combinations of these factors contribute to

^{1/} Ibid., p. 427.

explaining the resulting lack of direction provided for national policy on housing and income support programs. Legislators and federal program officials have become adept at using evaluation studies to buttress political stances, blunting some findings while vociferously stressing others to win support for favored programs. However, little work has been done on documenting the general impacts of such research on specific congressional votes for creating, maintaining, or cutting off program funding.

What is apparent is that a tentative concensus seems to exist among federal legislators (and even among many evaluators) to use rigorous, controlled social experimentation chiefly when major policy decisions are at stake. This generalization is not meant to imply that federal agencies on their own might not support, or have not supported, use of experimental methods in assessing internal program work; the Bonich 1/ listing clearly cites cases where such experiments have been carried out. The relatively small number of times that Congress and even federal agencies have funded such research, however, indicates that social experimentation is regarded as a special technique which produces its best results when applied in largely institutionalized or controlled social environments. A combination of design problems and the absence of sufficient political will to overcome ethical opposition to randomization of treatment and control group status, have given evaluation critics a strong foothold from which to

1/ (Ref.).

protest if not defeat the use of robust experimental techniques, unless sturdy policy-specific results can be pledged in advance.

Evaluability Assessment.

In order to deal with the skepticism and suspicion engendered by autonomous evaluations, some practitioners have developed evaluative methods aimed directly at usability by program managers and policy makers. The most sophisticated and comprehensive examples of these methods have been developed by Joseph Wholey and other staff members of the Urban Institute, and include a package of techniques, embracing evaluability assessment, rapid feedback evaluation, performance monitoring, and "intensive" evaluation. Other methods, notably rapid (or rural) reconnaissance analysis, similar to those employed by the Bank's Rural Operations Review and Support Unit (RORSU) ^{1/} and described in Chapter II, will also be discussed below.

Evaluability assessment appears to be a program review technique which seeks to determine whether program performance is likely to be adequate to reach its goals and whether a subsequent program evaluation is apt to be useful in improving performance. It is conducted by means of interviews in which program managers are asked for their views about implementation and objectives. The evaluators then seek the opinions of other groups on the program's course, examine program operations in the field, and provide a plan for formal assessment of actual priorities, effectiveness, and efficiency. If the

^{1/} RORSU has been renamed its department's Monitoring and Evaluation Unit.

plan is accepted, then it is used as the basis for actual evaluation activities.

It is not clear, however, what the initial and end points of an evaluability assessment are. It appears that the aim of the technique is to get program officers and managers to agree to evaluation procedures that they will regard as a sound basis for deciding upon changes in program operations.

Wholey and his colleagues outline eight iterative steps in the evaluability assessment method. First, the program under consideration is defined in terms of goals, processes, and interested parties. Next, opinions on program objectives, activities, and assumptions are gathered from the range of program information users. The assessors document each group's account of how inputs are supposed to yield results. The third step, a crucial one, is to set out the logical model reconciling these viewpoints in realistic terms. The fourth stage pins down the extent to which program objectives and information needs have been defined in measurable terms, specifying the ways assessors themselves would chart program progress. Next, operations in the field are observed to see if the assessment captures what the program is doing or is likely to do in reality. In the sixth phase, more modeling and analysis are done to synthesize viable means of measuring and enhancing the attainment of program goals. Here, "equivalency models," based on real operational performance are formulated. The seventh step identifies the actual options that could be taken to meet desired objectives. In effect it provides the final set of alternative work plans for the program. The final step is to present the assessment to

program managers and other chief information users, and obtain feedback on the assessment's adequacy as a program tool.

These evaluability assessment activities are fully diagrammed in the "path" or "flow" model shown in Figure 9.1. Each activity is linked by a causal connection between events (to be specified), as outlined in Figure 9.2. The example of a projected methadone treatment program is presented in Figures 9.3 - 9.5. The evaluability assessment for the last three figures documents the "logic" behind the program: it includes questions for program administrators and record-keepers, and sets out "objectively verifiable indicators" to be used in measuring actual program performance.

Wholey and his Urban Institute colleagues have been able to show federal program managers, through evaluability assessments, that only a portion of their programs were actually "measurable" in performance terms, and have gone on to recommend ways of gauging real program progress. The programs reviewed have included multi-million dollar efforts such as Community Mental Health Centers Program of the National Institute of Mental Health, and the Appalachian Regional Commission's (ARC) Health and Child Development Program. The evaluability assessment of the latter was completed in six months in 1976, at a cost of \$50,000 and led to ARC's decision "to monitor systematically the performance of all ARC health and child development projects and to identify and evaluate the effectiveness of "innovative" health and child development projects." 1/

1/ Wholey, 1979, pp. 35, 22-35, passim.

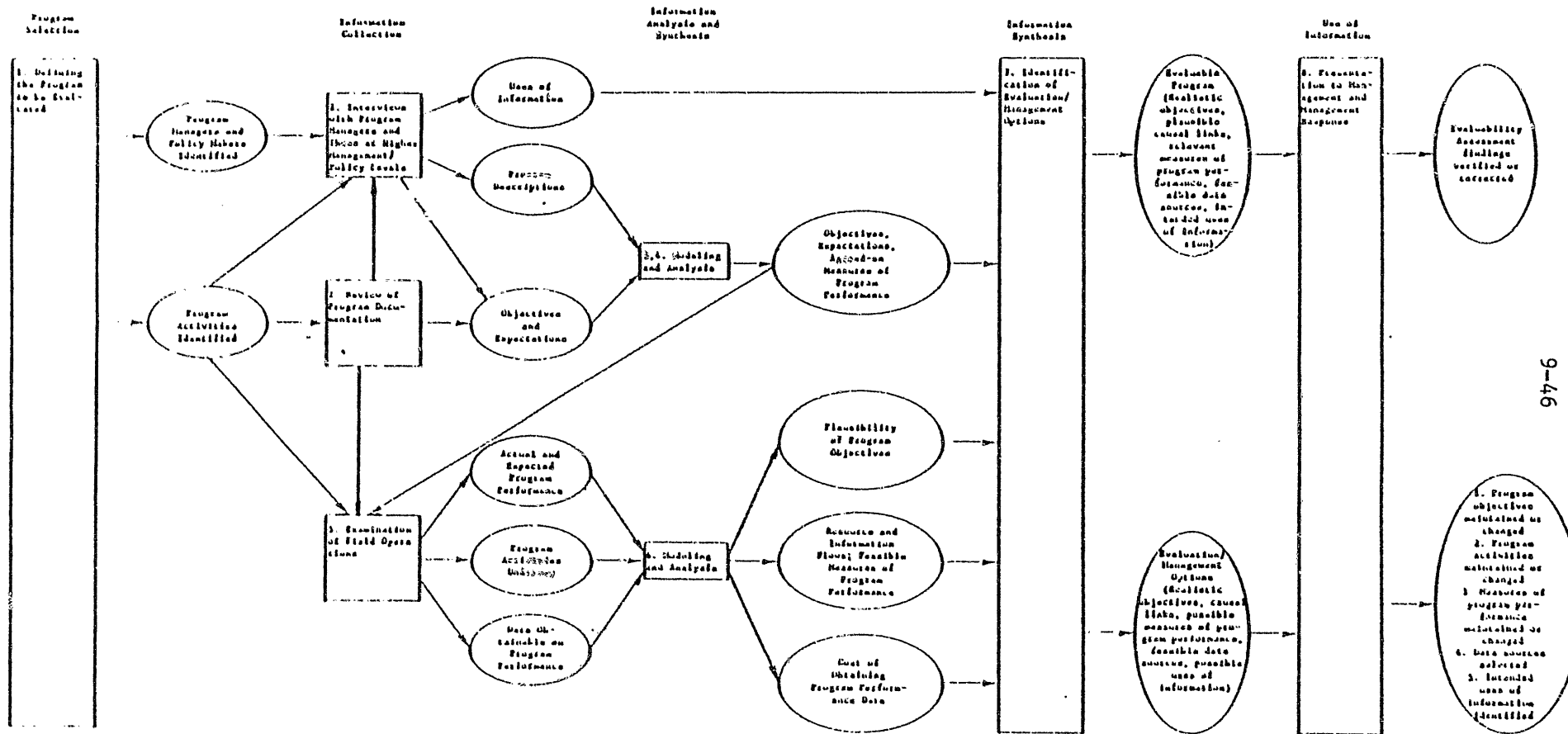


FIGURE 9.1: EVALUABILITY ASSESSMENT ACTIVITIES AND PRODUCTS

Source: Joseph S. Whaley, *Evaluation: Promise and Performance*, 1979, p. 51

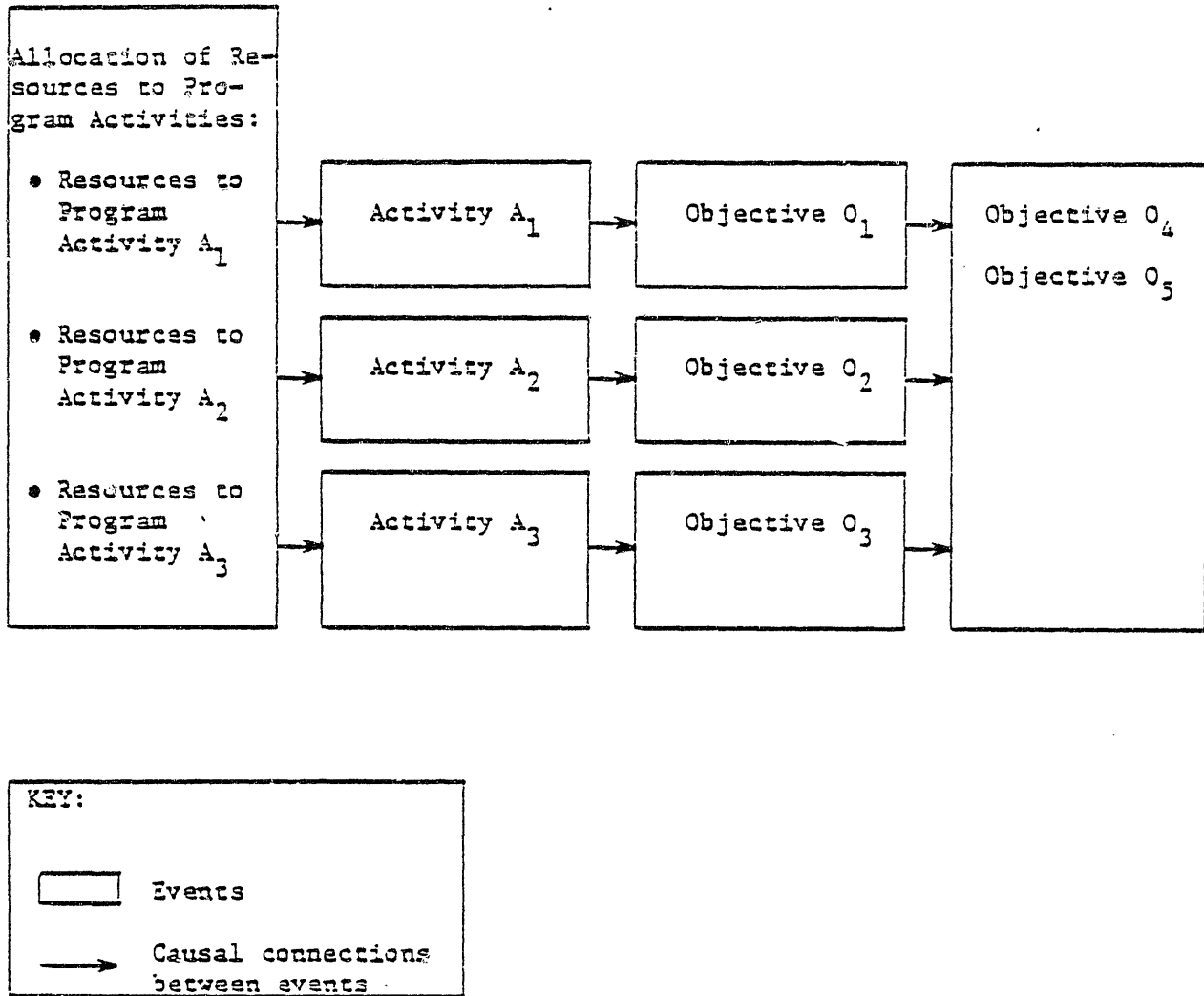
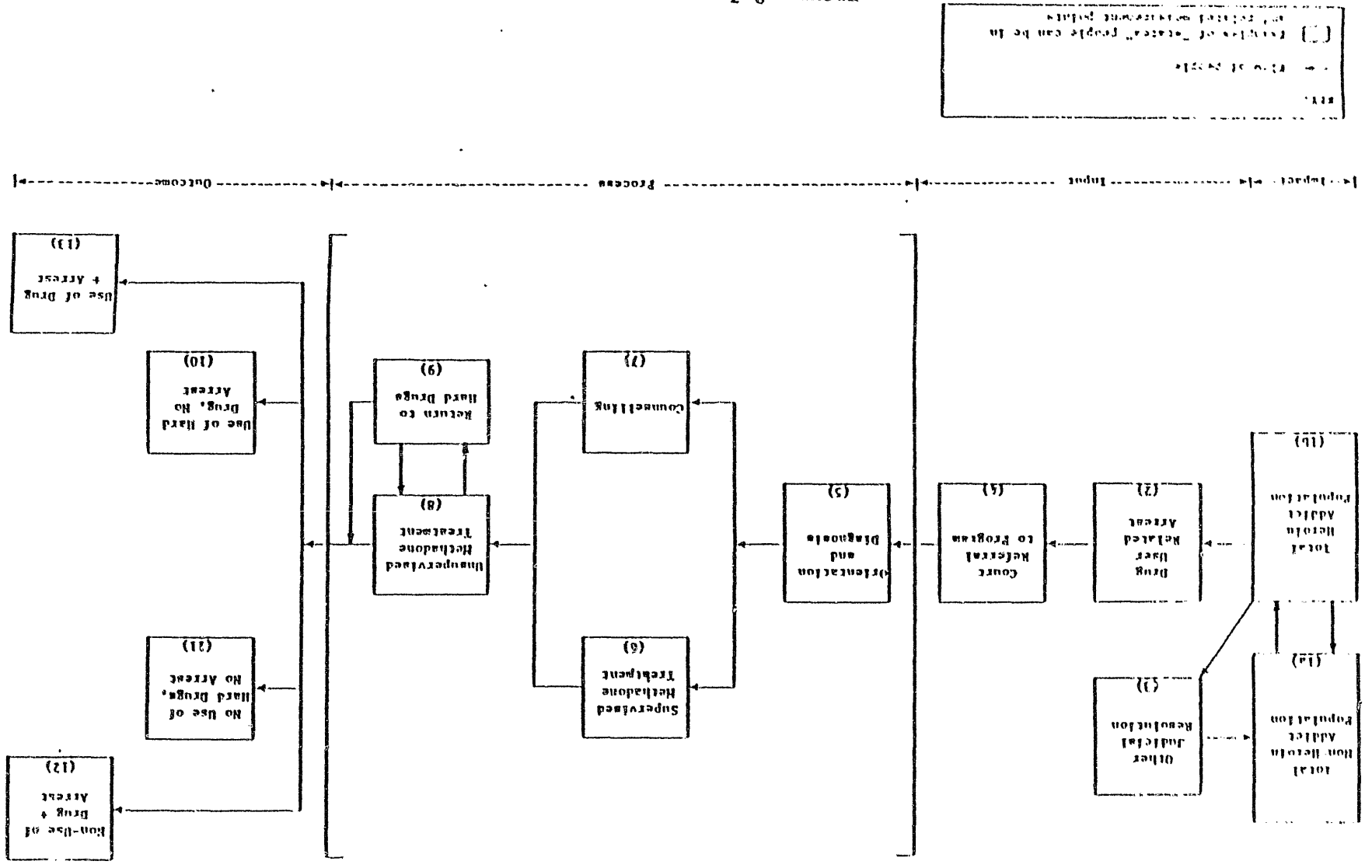


FIGURE 9.2 : LOGIC MODEL FOR A TYPICAL GOVERNMENT PROGRAM

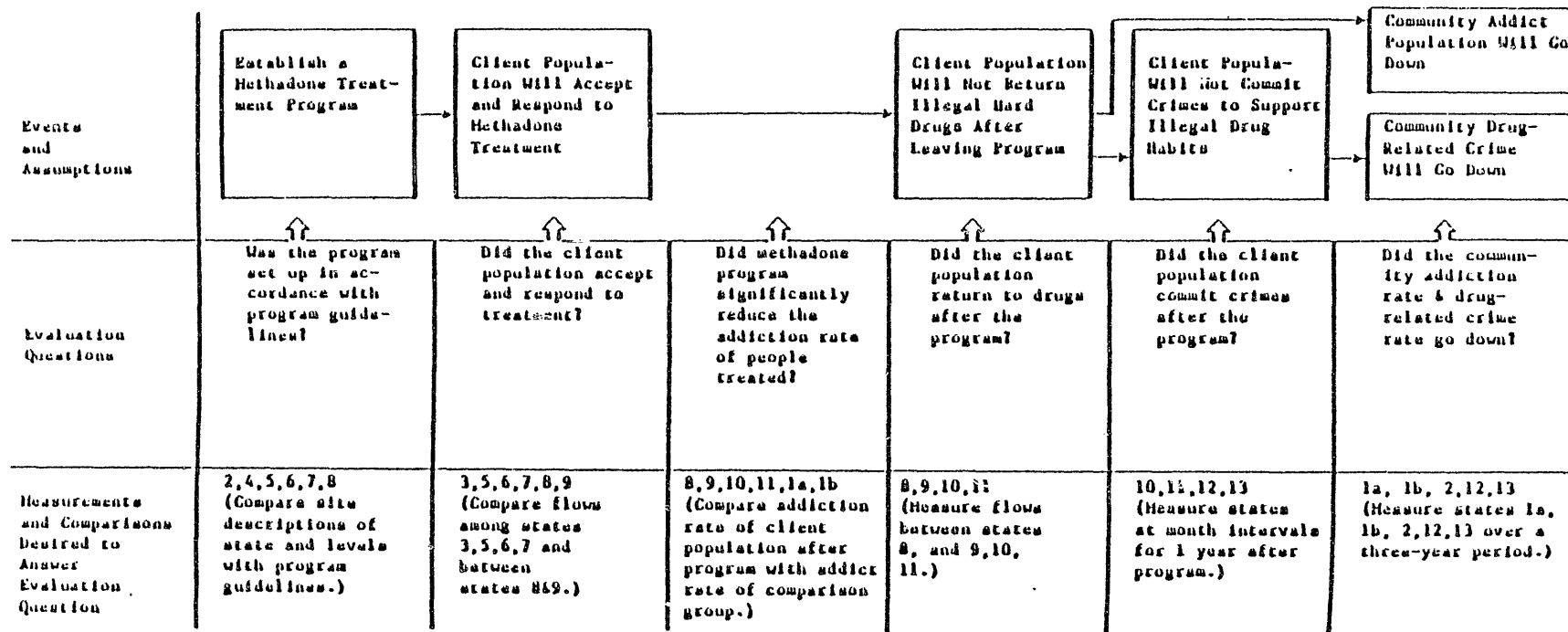
Source: Joseph L. Wroley, Evaluation: Promise and Performance 1977, p 59

Source: Lee H. May et al., The National Institute's Information Machine: A Case Study of the National
 Evaluation Program, Urban Institute, August 1977, Annex D, p. 15, cited
 in Johnson, et al., p. 25.

FIGURE 9.3 : MEASUREMENT MODEL FOR A METHADONE TREATMENT PROGRAM



EX: ... of people
 ... of "outcast" people can be in
 ... related measurement points



KEY:
 [] Event
 → Linking Assumption
 Numerals Measurement Point
 on Figure 9.4

FIGURE 9.4 : METHADONE TREATMENT PROGRAM: LOGIC EVALUATION QUESTIONS, MEASURES OF PROGRAM PERFORMANCE

Sources: Adapted from Joe N. Nay et al., The National Institute's Information Machine: A Case Study of the National Evaluation Program, Urban Institute Contract Report, August 1977, Annex D, p. 17, *et al.*
 in Joseph J. Wholey, *op. cit.*, p. 30.

FIGURE 9.5: MEASUREMENT MODEL FOR A METHADONE TREATMENT PROGRAM (continued)

Measurement Point	Definition of The State	Measures of Interest	Measurement and Data Collection System Available
1a, b	Total populations addicted and not addicted to heroin who were never in program	number of people	No
2	Addicts arrested for any offense	number	Yes
3	Arrested addicts who are not in program while under jurisdiction of court	number, rates to and from other states	Yes
4	Arrested addicts referred to program by court	number	Yes
5	Arrested addicts in orientation and diagnosis component	number, flows to other states	Yes
6	Clients in supervised methadone treatment	number, time in component	Yes
7	Clients in counseling component	number, time in component	Yes
8	Clients in unsupervised methadone treatment component	number, time in component, rates	Yes
9	Clients in program who return to using "hard" drugs	number, time on hard drugs	Yes
10, 11	Ex-clients using or not using hard drugs after the program who are never arrested	number, time in each state	No
12	Ex-clients arrested but not using hard drugs	number	Yes
13	Ex-clients arrested and using hard drugs	number	Yes

Source: Wiley, et al., p. 24.

The success of evaluability assessment seems to depend on whether program managers are really convinced that the data collected by evaluators will in fact chart the factors for which they are responsible; and whether they will then sponsor additional "performance monitoring," "rapid feedback evaluation," or "intensive evaluation" of the their program.

Wholey and his associates have carried out these "later stage" techniques on government programs. The first in the sequence, rapid feedback evaluation (RFE), aims at preliminary evaluation of a program's performance, and at outlining the steps toward a fuller-scale evaluation. RFE is an optional step between evaluability assessment and full or intensive evaluation.

Like evaluability assessment, RFE has begun to find an audience. In one HUD program, Operational Breakthrough, Congress requested data on the program's supposed advantages in creating industrialized housing. An RFE team collected documentation on program objectives and progress toward their achievement from HUD files, interviewed 73 staff members throughout the program, and observed four of the nine Operational Breakthrough sites. This research allowed the team to conclude that, while housing units were built and sold as planned, the program did not directly foster significant innovations in housing production technology, did not measurably affect the country's housing production, and failed to prompt savings and loan institutions to change their involvement in financing such housing. The research effort took six months (with ten person-months of work), and led to a

second stage, fuller evaluation, along the same lines designed for the RFE. Wholey notes that the two-stage evaluation provided information more rapidly and cheaply than the single evaluation that HUD had originally projected. 1/

The technique called "performance monitoring" by Wholey is defined as "periodic measurement toward program objectives," comparing actual with expected performance. Performance monitoring goes beyond RFE in establishing a full logical framework for the program and a reasonably complete program of inputs, outputs, and activities-including establishment of data sources, data collection, assessment, and presentation. USAID's "logical framework" provides an example of performance monitoring in which explicit standards of expected performance are stated at or prior to project appraisal and are subsequently measured on the basis of the Agency's own records and agreements with information users as to the events to be monitored, the measures to be used, and the intended uses of the research results. (See Table 9.2).

Assessment of User-Oriented Evaluative Techniques

Techniques such as evaluability assessment, RFE, and performance monitoring seem to be only as reliable and valid as the indicators or measures on which they are based. As to the usefulness of these techniques to their designated audiences, it appears that the techniques reflect the malleable phases of many federal programming objectives, for they seem to comprise little more than a last minute

1/ Ibid., pp. 89-90.

TABLE 9.2: AGENCY FOR INTERNATIONAL DEVELOPMENT'S "LOGICAL FRAMEWORK"*

	Summary	Objectively Verifiable Indicators	Data Sources	External Assumptions
Goal				
Project Purpose		End-of-Project Status		
Output		Output Targets		
Input		Budget People Schedule		

*Source: Adapted from Agency for International Development, Project Evaluation Guidelines, Third Edition, Washington, D.C., August 1974, p. 11.

Reproduced in Wholey, op cit., p. 128.

rehearsal or preliminary run-through of a program's progressive steps and their rationale. Whilst going over those considerations may be a valuable exercise for a program's implementers and information consumers, it is difficult to see why such a review is not a required part of program appraisal and implementation processes. Development planners especially might reasonably expect that the eight iterative steps in evaluability assessment would be reviewed before completion of the design process. "Evaluability assessment" itself seems to be a misnomer. The content of the activity seems to be a mediation between program implementers and policy makers on questions of what they expect in the way of program results. In short, "evaluability assessment" seems merely to describe and recommend as a special enterprise what many practitioners regard as normal project appraisal or preparation activities.

This formal objection aside, the three techniques do seem to have been relatively useful to some program managers in clarifying measurements of program performance and indicating possible program adjustments. The suspicion remains, however, that the techniques are little more than commonsensical extensions of the assumptions and procedures according to which program development is to be charted. Although each of the review functions may help give managers additional perspective, the techniques by their checklist nature make no contribution to the social and economic theory on which a program has been based (this issue may have received no other consideration either). More demanding audiences may not be satisfied to know only the course a

project has taken; they may also want to know why particular outcomes occurred. For example, although the RFE review of Operation Break-through yielded interesting quantitative results concerning changes in industrialized housing production, the technique could in no way account to HUD for the reasons why those housing production results were (or, in this instance, were not) obtained. HUD was thus prevented from usefully generalizing the program's outcomes by the lack of scope and depth in the RFE process. Despite the claim of cost-effectiveness for this RFE exercise, its very rapidity shortchanged the specific audience seeking lessons on the benefits or general effectiveness of program policy that might have been derivable from a more developed technique. Tailoring an evaluation to an audience does not guarantee that its results will be those the audience requires.

The same shortcomings can be identified in "performance monitoring", in the abbreviated form discussed by Wholey and implemented by such federal institutions as AID. The laying out of a framework for verifying program progress according to "objectively identifiable indicators" does nothing to demonstrate why that framework is better than other possibilities. Such a demonstration can only be made with the formulation of a theory explaining why the chosen indicators are the most significant, something neither Wholey, nor DAI, nor even the World Bank's Operations Evaluation Department have set forth in their assessment procedures. As Wholey is forced to admit (along with other consultants reviewing USAID's in-course evaluation guidelines), "At AID, they do not have 'hard' evidence that this

monitoring system is effective. AID evaluators have apparently observed that while mission reviews of individual projects formerly consisted of debates over project inputs, they are now devoted to discussions of projects inputs and outputs." 1/

The performance monitoring technique, in short, carries with it little or no justification for the claim that it reviews program progress for its intended audiences better than any other method.

Wholey's entire discussion of evaluability assessment, rapid feedback evaluation, and performance monitoring finally avoids considering squarely where and how each of these techniques blends with the others. He does not propose criteria for recognizing when the indicators used in evaluability assessment, for example, have to be amended before evaluation activity proceeds to later stages. Wholey, basing his claims for his evaluative approach on its timely utility to user audiences, actually groups the conventional program review questions under new rubrics at different stages of program planning and implementation, and makes little argument for why any particular approach might be primary, other than citing managers' need to know the details of a program's course. It would be interesting to investigate precisely why the federal managers employing Wholey's techniques did so in the first place. It may well have been that programs (or their managers) were so new in their respective agencies that it was felt that any evaluation method would help gauge program success. Wholey, a former observer of and practitioner within the federal government may

1/ Ibid., p. 129.

simply have offered the easiest evaluation package from a marketing standpoint.

Interim Assessment of the DEDRB Evaluation vis-a-vis Trend II

Like the Wholey approach to evaluation, the DEDRB study committed resources to providing timely, relevant information to project managers to assist them in solving problems of project implementation. A formal design was devised to plan the evaluation system. It provided for negotiations between policy makers, program managers, and researchers to define program objectives and the key issues in the research framework. The design was made iterative to include the capacity for modification (see Figure 9.4). Plans were also made for stating and meeting the information requirements of each stage of project development, targetting the kinds of information most likely to be demanded by the various evaluation audiences (see Figures 9.5 and 9.6). ^{1/}

The advantages of the DEDRB design over the Wholey approach derives less from the greater scope and specificity of the information provided to research consumers than from the modeling of the design in terms relevant to the urban housing sector generally. DEDRB evaluation studies have been aimed not only at answering managers' questions about progress in meeting project objectives, but also at formulating findings

^{1/} These figures are discussed at greater length in other DEDRB publications on evaluation systems. See Michael Bamberger, Planning and Evaluation System for an Urban Shelter Program: Key Issues for Program Managers; Methodology for the Evaluation of Urban Shelter Programs: An Overview; and A Basic Evaluation Package, volumes 1-5, of the DEDRB Evaluation Series, DEDRB, 1980-1981.

FIGURE 9.6

A LIST OF POVERTY INDICATORS
SUGGESTED DURING A MEETING AT KECAMATAN WCNO SALAM*

- | |
|---|
| 1. A man is poor if he has no land, or if he has to rent out his land. |
| 2. A man is poor if his house costs less than 10,000 rupiah (± \$16) to build and nobody would want to buy it, and if anybody bought the land the house is on they would rather burn the house than use it; an exception is persons who are poor themselves but who have inherited old teak houses. |
| 3. A man is poor if he has to work as a buruh (paid laborer). (A man is rich if he can hire buruh). |
| 4. A man is poor if he has to make handcrafts, i.e., if he cannot live by farming alone. |
| 5. A man is poor if he has to borrow a bicycle when he wants to go someplace. |
| 6. A man is poor if he has to mix his rice with cassava or corn. |

*Source: Ann Soetoro, "Prosperity Indicators for Java," Washington, D. C.: Development Alternatives, Inc., March 1979.

and conclusions in terms of testable hypotheses that could be of theoretical and practical use to housing sector planners elsewhere. In other words, the evaluation format took into account the need to show how and why particular outcomes occurred and to incorporate these "appreciations" into observations on project performance and design, at the same time that it attempted to meet the immediate information needs of the Bank. Hence, for example, the evaluation produced hypotheses and findings on housing affordability and accessibility, on inter-household income transfers, and on implementation efficiency that are largely testable. Wholey's "evaluation products," as outlined by him, do not include studies formulated so as to be replicable.

This is not to say, however, that all the audiences concerned were equally satisfied with DEDRB evaluation results. At times the statistical rigor of some studies required lengthy data collection, processing, and analysis, causing impatience on the part of evaluation consumers and raising the question of the findings' possible "operational relevance." It is now better understood that there are costs in time associated with careful measurement of impacts. The level of methodological validity in many studies compensates for initial costs in time by reducing the costs of evaluation studies in subsequent Bank-supported shelter programs. Despite substantial costs, there are also benefits to be derived from the rigorous modeling of urban shelter development in terms of providing a basis for allocation of scarce resources in developing countries.

The Third Trend: Indicator Use and Refinement

As noted in the preceding analyses, modeling of evaluation hypotheses to suit program environments, and clarification of the uses and users of evaluation results have both involved refinement of the indicators and measures used in establishing evaluation systems. This development in itself constitutes the third trend in evaluation practice in recent years. Depending upon the explicitness of the theories put forward to defend the use of certain measurements, and the rigor of the indicators' statistical manipulations, evaluators have been able to clarify to differing extents the significance of particular indicators to evaluation methods. The trend has included questioning and re-analysis of evaluation results as well as adaptation of particular evaluation procedures to new settings and to accommodate various social program objectives. In a related development, measurement refinement has led to a re-evaluation of evaluation itself. This minor trend, called "metaevaluation," will also be treated briefly below.

Although modification, substitution, and re-interpretation of indicators has been carried out across a wide variety of program sectors, including housing, education, employment generation, rural development, etc., not all refinements have been accomplished in the same manner, or with the same degree of cogency. The trend has essentially taken two different forms. The first is to propounding the use of indicators themselves, as a convenient shorthand for charting processes of economic and social change via evaluation studies, particularly in relatively brief evaluation exercises. Such assessments

include the rapid feedback evaluation technique, performance monitoring, and evaluability assessment analyzed above. But the most forceful advocates of the indicator technique have been proponents of rapid (or rural) reconnaissance analysis (RRA). Details and examples of this method will be presented below.

The second form which Trend III has taken involves the specific, rigorous assessment of the use of particular indicators, as part of a formal evaluation design. Much of the scholarly literature devoted to evaluation research has been taken up with documenting progress, or expounding arguments for or against particular devices in this aspect of the third trend. Only a few brief examples of strict indicator re-assessment can be presented here. They will be taken from the areas of evaluation research in which most formal developments have occurred, namely, investigations of educational development; employment creation and training efforts; and income support programs.

Rapid Reconnaissance Analysis

In relatively short-term, time-pressured evaluations, researchers have frequently been asked to find indicators quickly that will capture the extent of project inputs, outputs, and first-run effects. In the fiscal period, 1976-1980, the World Bank's agricultural and rural development department, through RORSU, was called upon to see to it that such systems were incorporated in several projects, including agricultural extension and nutrition projects. Even more emphatic advocates of the technique have been firms such as Development Alternatives, Inc. (DAI), as well as particular evaluation practitioners

within USAID and other development agencies, who have labeled their method rapid (or rural) reconnaissance analysis (RRA). RRA consists of compiling data on a set of indicators deemed accurately illustrative of progress towards realization of particular objectives. Data are typically collected in a 3 to 6 week period during or soon after project implementation. Given budget and time constraints, RRA defenders claim that data collection on proxies for phenomena which are in themselves more difficult to observe and capture directly is preferable to no data collection at all. They argue, moreover, that gathering these data iteratively will not only reveal the sorts of progress achieved by the project, but may help also to uncover the sources of observed changes.

Practitioners of RRA have pursued the use and refinement of different development indicators to varying degrees. Some attempt, for example, to elicit indigenous criteria for measuring a common development concept such as poverty, and to establish a hierarchy of productive and cultural factors symbolizing little wealth, less wealth, no wealth, etc. ^{1/} As represented in Table 9.3 below, Soetoro produced characteristics intended to represent low, medium, and high levels of prosperity for a variety of common household features. Similarly, she computed statements on the value of land, labor, housing, food, etc., the scarcity or absence of which were considered to serve as gauges of rural peoples' economic status. (See Figure 9.6). This approach seems to be of potentially greater use than many others, in that it clarifies the terms and referents of poverty along a more graduated, culturally

^{1/} Ann Soetore in Honadle, 1979: 11-14; Conlon and Wiggins, 1980.

TABLE 9.3

MEASURES OF PROSPERITY IN RURAL JAVA*

INDICATOR	PROSPERITY LEVEL		
	LOW	MEDIUM	HIGH
House	Bamboo	Combination	Brick and Plaster; Teak
Rooms	1-2; small	--	Many, large
Floor	Dirt	Bricks covered with cement; limestone blocks	Polished cement block
Roof	Straw; fronds	--	Tiles
Windows	None	Wooden with slats	Wooden frames with glass panels
Bedding	Mats on floor	Bamboo slat beds with mats	Wooden or iron beds w mattresses and mosqui nets
Lighting	Small oil lamps	Hanging kerosene lamps	Home generator
Water Source	Neighbor's well; river; spring	--	Own well
Toilet	Outdoor not enclosed	Outdoor enclosed	Indoor
Transportation	None	Bicycle; draft cart	Motorcycle; scooter; truck; minivan
Entertainment Equipment	None	Radio, Tape recorder	Battery TV
Refreshment served to interviewer	None; tea without sugar	Tea with sugar; other sweet drink	Tea or coffee with sugar plus snacks.

*Source: Adapted from Ann Soetoro, "Prosperity Indicators for Java," Washington, D. C., Development Alternatives, Inc., March 1979.

relevant scale than the simple use of "presence or absence" of television, radio, wrist-watch, etc., as a proxy for the same concept.

In general, RRA advocates seldom propose a theory to make chosen indicators convincing, beyond appeals to "common sense," utility, readiness for collection, etc. Hence, the indicators proposed may be more or less inaccurate standards for measuring steps toward economic or social development. In one USAID rural road project evaluation, conducted over several weeks, the presence of tin roofs was being used as a sign of economic well-being, until the AID team learned that Liberian villagers often went heavily into debt as a result of installing such roofs. ^{1/} In a DAI publication advocating the user of elephants as an indicator of illegal logging in forested areas in of Central Asia, there is no notation of the fact that elephants may also be used for legal goods portage, personal transportation, religious functions, and even as a store of value. ^{2/} Although DAI staff go on to note that indicators should be assessed in the light of actual behavior in a development project area, they again stop short of proposing an explicit theoretical frame of reference to account for indicator validity or the occurrence of observed phenomena. ^{3/} As noted in the earlier critique of RORSU, indicators cannot explain why events take place; agricultural yield data, for example, at best produce

^{1/} Impact on Rural Roads in Liberia, USAID Impact Evaluation Report No. 6, June 1980, Methodology Section, p. 3.

^{2/} George Hondale, "Rapid Reconnaissance Approaches to Organizational Analysis for Development Administration," Development Alternatives, Inc., December, 1979, p. 8.

^{3/} Ibid., pp. 17-18.

representative statistics on the achievement of output objectives, but cannot by themselves show how or why a certain output has been achieved.

The favoring of particular indicators tends to register the wholesale presence or absence of a certain phenomenon, whilst it is the relative proportion of that phenomenon's occurrence that may be the real gauge of social or economic change in project areas. In the AID-Liberian Rural Road Project Evaluation cited above, the investigating team was able to sort out short-run effects from likely longer-term results. However, it failed even to comment on specific methods for weighing long-run against short-run or positive against negative effects.

It was noted, for example, that the project had contributed to a reduction of transportation costs in the area, increased marketing activities and increases in school attendance and the use of health clinics; at the same time, the project was deemed to have negative short-run effects, including traffic deaths, and the higher incidence of water-borne diseases due to the increased number of stagnant pools created along the roadways. Over the long term, negative impacts were expected to outweigh positive ones because the presence of the roads was also expected to increase land values in the area and thus to attract wealthy, city-dwelling purchasers, aggravating both the socio-economic differentiation in the vicinity, as well as the distribution of power between rural and urban zones. These undesirable consequences were believed to exceed the positive ones of opening up at least more land

for subsistence and commercial agricultural activities. 1/ However, no means were provided for actually squaring the costs and benefits of each development against the others. Further, quantitative data to complement the qualitative documentation of the effects reported to have occurred were severely lacking. 2/

If employed too loosely, data on indicators alone--as opposed to substantive data on economic or social processes or theories--are apt on the one hand to be frequently inconclusive or misleading, and hence to cast doubt on the value of indicators as an element in a comprehensive strategy for explaining the results of development projects. On the other hand, they have the equally negative potential of becoming ends in themselves. Projects managers and even some evaluators may be so oriented to goal-attainment studies that they sacrifice understanding about how project objectives are achieved for recognition of the achievements alone. 3/

In short, carried too far, the reliance on development indicators alone can draw attention away from the serious examination of the assumptions on which projects were based in the first place. Because institutions like USAID and the World Bank tend to place too low

1/ USAID-Liberia Rural Road Project Report, op. cit., Introduction.

2/ Though it may be defensible not have succeeded, within a given period, in evaluating positive and negative effects within a consistent benefit-cost framework, it is hardly defensible not to have tried. The techniques, including those for applying social weights to benefits and costs experienced by different groups, are available; and they have been successfully applied, for example, within the DEDRB program.

3/ This serves to compromise severely the utility, for example, of the World Bank's project completion reporting.

a premium on specifying the rationale behind the translation of project inputs into outputs, effects and impacts" it is often difficult for evaluators to account convincingly for the results of project implementation: certainly with respect to how they happened, and sometimes, even with respect to their very occurrence. Projects are rarely viewed as formal or informal experiments for the testing of economic, financial, or social hypotheses. Instead, they are typically treated as if the speculative thinking behind the experiment had already been proved correct.

There appear to be three features of rapid reconnaissance analysis which combine to account for these shortcomings. It has not been stated, but will be noticed that this approach is typically one of ex post analysis; that is, the evaluation is typically designed at the completion of project implementation, rather than during project preparation. This helps to explain, at the same time that it accentuates the deleterious effects of the failure to build the evaluation on firm theoretical foundations, and to maximize opportunities throughout the project's life to quantify the key relationships. An uncharitable view is that such evaluation is eclectic and opportunistic, and relies entirely for any success it may have on the skill, ingenuity and persuasiveness of the evaluators involved, rather than on any advances in technical knowledge.

Re-assessments of Exemplary Indicators in Specific Social
Program Evaluations

This section will review methodological implications of the evaluations of three American social programs, relative to the import of the DEDRB studies on housing affordability and income supports among related households.

In 1975, Thomas Cook and his associates re-assessed Samuel Ball's and Gerry Bogatz's evaluations of Sesame Street viewing, performed in 1970 and 1971. Ball and Bogatz found in their pre-test/post-test controlled experiment that encouragement of children to watch the program by paid professionals using promotional games, books, and toys, resulted in both increased viewing and learning. They also maintained that the effects of viewing alone were associated with clear learning gains. Cook and his associates confirmed the finding that promotional encouragement aided viewing and learning together, but found that most of the effects of viewing alone on learning were not statistically significant.

Through a number of multiple regression and covariance analyses, Cook and his colleagues went on to demonstrate that Ball and Bogatz had inadequately accounted for pre-test differences between "encouraged" and "non-encouraged" or treatment and control groups. Ball and Bogatz had calculated the differences between the pre-test and post-test scores for both groups, and allowed the difference in gains to stand for the program effects, whereas Cook's team showed that such a manipulation cannot attribute general gains achieved to viewing Sesame Street alone. Individual maturation, rather than viewing itself might

have caused the results. The secondary evaluation builds on the primary one, however, by manipulating the same sets of raw data. Cook and his associates did not have to replicate the experiment in order to achieve their more sophisticated results with regression analyses. That development in itself reflects an advance in the state of the art of evaluation research during the 1970s.

In a similar exercise, a group of professional evaluators at Abt Associates and Northwestern University re-examined in 1978 the primary results of a 1974 Rand Corporation assessment of the Education Voucher Demonstration in Alum Rock Union Elementary School District. Under the voucher concept, parents choose a school for their children within the district and receive a voucher or credit equal to the cost of the childrens' education; this sum is then credited to the school upon presentation of the voucher at enrollment. The assumption is that the competitive bidding for particular schools thus enhances educational quality by making schools more responsive to childrens' needs. In a comparison of voucher to non-voucher schools in the first year of the demonstration program, based on reading test mean scores, the Rand analysis found that performance among primary graders declined. The decline, while small, was registered repeatedly after use of a variety of measures as controls for socio-economic status differences.

The Northwestern-Abt group was able to obtain individual students' reading test scores, prior to and during the demonstration period, for a part of the Rand sample. The added data made possible the creation of a quasi-experimental design involving multiple test series

and a more refined unit of analysis, the individual rather than the mean test score. At the same time, the smaller sample size limited the generalizability of the findings. The re-evaluation indicates that the drop in reading scores was confined to a few programs within particular schools which featured non-traditional educational curricula. Overall, compared to the non-voucher program, the voucher group taking traditional curricula registered no or only slight deleterious effects. Although different data sets were used in the Rand and secondary studies, thereby raising the costs and increasing the length of time involved in conducting a "new" analysis, the more elaborate evaluation methodology used in the Northwestern Abt study appears to have justified the extra effort involved, especially because major policy decisions concerning the future of voucher programs were likely to be drawn on the basis of the evaluation.

What this experience seems to illustrate more than anything is the perils of inadequate theorizing before designing programs, whatever the evaluation techniques. Economic theory would not necessarily predict, for example, that such a voucher system would lead unambiguously to improve reading text sources in all participating schools. If one were to construct a set of hypotheses based upon Tibout's hypothesis, ^{1/} one would probably predict that there would be specialization among schools, with some coming to excel, for example, in reading; others, perhaps in athletics. Furthermore, it is unlikely that major changes would occur within a short interval.

^{1/} Ref.

As discussed earlier, the New Jersey Negative Income Tax Experiment (NJNIITE) represented a major attempt to base public policy on the results of social experimentation. Beginning in 1968, the four-year program measured the effects on labor response ("L") of a guaranteed benefit (or income support) level ("g") and the marginal tax rate ("r") assessed at that level among work eligible male-headed families, $L = f(g,r)$. As such, the crucial variables to be gauged in the response equation were pre-experimental earnings, income, wage rates, and hours worked, as well as the calculation of earnings, wage rates, hours worked, and other income while the program was being put into effect. However, changes in implementation procedures of the NJNIITE, relative to other support programs for which local families could apply, resulted in measures being taken which were different from those established at the baseline.

As Rossi and Lyall note in their critique of the NJNIITE evaluation design, checks were not made consistently to ensure that both earnings and non-labor income were each fully and independently reported. Wage rates were also calculated on the basis of reported gross earnings and hours of work, rather than disaggregated on the basis of straight time versus overtime, differentials between shifts, etc. The result is that the values attributed to L, g, and r for different families or program sites were comprised of weighted averages of coefficients which often could not be broken down into the actual values of the variables measured. This faulty data reporting hindered the potential for more exact interpretation of the results, restricting not

only the generalizability of the findings, but also the confidence of policy makers in the direct results based on these data and procedures. Because these procedures were amply criticized following the first rounds of result reporting for the NJNITE, subsequent debate and efforts in other NITEs "have made economists generally more sensitive to the need for improved survey techniques and primary data collection." ^{1/}

Taken together, these three cases reveal a number of things about the state of the art in indicator refinement during the late 1970s, concurrent with the DEDRB evaluation. First, scholars have increasingly taken to presenting their data so as to facilitate their reinterpretation. This requires that sources of data, design formulation, indicator selection, and analytical procedures be sufficiently detailed to permit evaluators to disassemble the research plan, review the adequacy of the component parts, reformulate and retest the original calculations. Through this "sharing" procedure, practitioners were not only exposed to a greater number of models and statistical techniques for validating research results, but could also specify ways to augment and more fully exploit the original data base. All three "secondary" evaluations cited in fact questioned the manner in which the original data sources were determined, turning to other data bases or making comparisons to other bases which might have enriched the primary efforts.

Second, the general trend in indicator refinement has been to limit the conditions of validity or generalizability attached to the

^{1/} Rossi and Lyall, 1976.

primary results. During the past decade, researchers have improved and specified the tests of internal, external, statistical conclusion, and construct validity that rigorous social experimentation must be able to meet. 1/ In the Sesame Street research, the threat of the maturation factor to internal validity of the testing was not taken sufficiently into account. In the Voucher Experiment, the question of measurement reliability, a threat to the design's statistical conclusion validity was better addressed by the disaggregated student reading scores than by the school averages. In the review of the NJNITE, a host of factors was cited which threatened the experiment's internal, external, and statistical conclusion validity. These included factors of instrumentation, local history, the reliability of measures, the interaction of multiple treatments, selection-treatment interactions, etc.

Replication of testing for the results of watching Sesame Street has hence required wariness of maturation and external validity factors in establishing the causes of the effects observed. 2/ Proponents and antagonists of education voucher programs have likewise had to sift through the results of testing in different areas to learn the extent to which the voucher system was likely to have delivered the educational benefits promised in the competition between school

1/ Cf. Campbell and Reichardt, 1979, introduction; Cook and Campbell, 1979.

2/ Diaz-Guerrero, 1974, 1976.

curricula and formats. ^{1/} Finally, as noted above, the NITEs subsequent to the initial New Jersey and Pennsylvania programs did attempt to meet the threats to validity posed by the original program design by using more refined measures of household structure, income, labor time, site-specific characteristics, interaction with other programs, etc. ^{2/} Given the heightened official and public awareness of the issues of social experimentation's reliability and applicability, evaluators have become increasingly guarded in the claims they make for the validity and value of evaluation exercises.

Despite advances in the refinement of measures and indicators (culminating in the proposal of 35 factors to account for every level of validity) evaluation methodologists continue to face certain problems in the application of statistical techniques to social program evaluation. These problems have been especially troublesome in repeated time-series design and some quasi-experiments. The problems include: (1) tests of significance which are frequently confounded by systematic cycles in data collected. To date, all methods of removing the cycles, including "transfer functions" and "moving average" models, ^{3/} tend to underadjust effects. Similarly, (2) removing seasonal trends remains a hindrance in measuring strict program impacts. Methods for separating month-to-month changes from coincident program changes cannot be counted as purely seasonal, and series have to be split at these points of change to calculate seasonal patterns. Hence, the parts of the series

^{1/} Wortman, Reichardt, and St. Pierre, 1978, pp. 210-212.

^{2/} Rossi and Lyall, *op. cit.*, pp. 415-416.

^{3/} Cf. Campbell, 1979.

just before and just after program initiation become series ends, but corrections for these parts are poorer than for mid-series points. Finally, (3) program changes themselves introduce a series of measurement problems: gradually introduced changes are usually impossible to uncover by this design, and where programs are initiated in response to an acute problem, which appears as a sudden change in a social indicator, ameliorative effects of the program can easily be confounded with regression artifacts.

Indicator Refinement in the DEDRB Evaluation and the
Academic Evaluation Literature

As recounted in the preceding chapters, DEDRB efforts have resulted in positive contributions to the literature, e.g., on affordability calculations, on determination of poor households' income sources, on hedonic indices of housing quality, and on statistical calculation of project effects through longitudinal studies. ^{1/} It has been demonstrated that a measure of total income rather than earnings income is a more illustrative, and probably a more reliable basis on income is a more illustrative, and probably a more reliable basis on

^{1/} These contributions include the following publications in the DEDRB Evaluation Series, 1980: Michael Bamberger, Statistical Evaluation of Project Impact Through Longitudinal Studies: Their Application in the Evaluation of the Impact of Urban Shelter Programs in Developing Countries; Dani Kaufmann, "A Theory and Evidence on a Strategy for Survival Among the Urban Poor;" David Lindauer and Dani Kaufmann, "Basic Needs, Interhousehold Transfers and the Extended Family;" Douglas H. Keare and Emmanuel Jimenez, "Affordability Income, and Housing Consumption;" Emmanuel Jimenez, "The Economics of Self-Help Housing: Theory and Some Evidence" and Emmanuel Jimenez, "The Value of Squatter Dwellings in Developing Countries."

which to assess households' ability to pay for low-cost housing. It was also learned that a regular, stable, significant component of this total income is the transfer income received from members of the extended family network of the participating low-income households.

As for hedonic indices of housing value, DEDRB studies have shown that determinants of the value of squatter dwellings are similar to those which apply to conventional "formal sector" dwellings, with the substantive exceptions of the value of sanitary facilities and the positive correlation of age with value in the Philippines case studied.

With regard to the statistical assessment of project impacts through longitudinal investigations, the DEDRB evaluation has specifically devised measures to maintain the internal validity of urban shelter program research. These techniques include clearer definitions of selected categories and devices to keep appropriate samples matched, such as recording full names in the data files, building redundancy into the coding, and devising logical consistency checks at the stage of findings analysis, all of which were aimed at offsetting the effects of substantial mobility among households. Statistical tests for measuring change in panel samples with nominal, ordinal, and interval scale variables were similarly devised or adapted for other contingencies of urban shelter projects. These included careful accounting for such crucial indices as changes in housing quality, in the earned income of household heads and complete households, in the number of households occupying particular structures, in food expenditures, and in job stability. Because the evaluation compared the project populations with

non-equivalent control groups, it was also necessary to devise ways to control for the differential effects of intervening variables through the use of partial correlation, cross-breaks, and multiple regression analysis.

Still other issues of measurement refinement confronted DEDRB evaluators, especially sample representativeness and accounting for "treatment" modifications via in-course project changes. The question of sample representativeness arose when results had to be generalized to the total low-income population to ascertain city-wide project effects. The problem was met by restricting conclusions to specific urban subgroups; much caution was needed when generalizing beyond those groups. Maintenance of the evaluation's methodological rigor was also made more difficult by the frequently changing pace of project implementation and the limited local research unit capabilities. The research design was amended often to cope with the rescheduling of specific projects' implementation, in order to provide analysis of critical operational issues at the same time that data sets were being assembled to measure project effects. Overall, the DEDRB statistical procedures appear to be fully concurrent with other technical adjustments discussed in the most recent academic evaluation literature.

Meta-evaluation

The final sub-trend, "meta-evaluation," expands the concept of measurement and indicator refinement. In essence, meta-evaluation denotes any re-analysis of an evaluation, from a simple qualitative critique to simultaneous, multiple, independent replications of the data

sets collected and analyzed as part of a "primary" evaluation. Meta-evaluation, then, can be considered as coincident with the development of evaluation research as a discipline.

Two complementary approaches have so far been taken to systematizing meta-evaluation activities. One, associated with the work of Thomas Cook and others at Northwestern University, involves categorizing the kinds of re-analysis undertaken by evaluators on the bases of (1) employing single or multiple data sets, (2) manipulating the same data sets used in the primary exercise or introducing additional ones, and (3) carrying out the meta-evaluation simultaneously with or subsequent to the primary research work. The aim of Cook's approach is to lay out as formally and logically as possible the steps taken to assess program effects by means of evaluation, then to test whether these steps were technically well-executed, and to review the relations between independent, intervening, and dependent variables. Where threats to validity are found, to redevise the evaluation structure becomes the chief aim.

The second approach, associated with the work of Michael Scriven, Gene V. Glass, and other "systems oriented" evaluation practitioners, starts with the relation between the evaluation and the program under scrutiny. From organizational, theoretical, and methodological viewpoints, it reviews how the evaluation attempts to assess whether a program will meet its objectives, and tried to project the degree to which specific objectives will be fulfilled. Part of Scriven's approach is perfectly compatible with Cook's rigorous

recalculation of formal evaluative procedures. The rest attempts to embrace the full rational-empiricist exercise of determining a program's value in all its economic, social, psychological, operational, and organizational manifestations. Scriven eschews no technique that may aid in revealing a program's potential. In addition to formal experimental design, his tack incorporates goals analysis, composite outcome measures, etc. Yet Scriven does not tacitly recommend an indefinite regression of evaluations, nor does he deny that time and cost figure in the value and performance of an evaluation. He does maintain that the meta-evaluation debate ought to be focused on learning from policy makers what appropriate goals for a program should be (or should have been), and that a reasonable selection and weighting of performance criteria are necessary in order to identify the best form of evaluation in each case. 1/

Conclusions

This book itself is a contribution to meta-evaluation studies much in the style advocated by Scriven and Glass. The manuals produced by DEDRB in its Evaluation Series derive from the approach propounded by Thomas Cook and his associates at Northwestern. Together they make a contribution to evaluation research per se and to nascent efforts at establishing a frame and field of reference for meta-evaluation.

This study has reviewed the decisions within the World Bank that led to the original form of the evaluation, and to its relationship with the implementing project agencies within and outside the Bank. On

1/ Glass, 1980.

balance, our judgment of the DEDRB evaluation, our "meta-evaluative" judgment, is that the benefits of the program have considerably outweighed its substantial costs. Questions as to these benefits, posed to the evaluators by the Executive Board of the Bank, have been answered with rigorously supported conclusions regarding the workability of the progressive development model for urban shelter programs. Some of the institutional conditions for program replicability have been investigated, and a multi-dimensional model for implementing new shelter project evaluations has been developed. And as already mentioned, contributions have been made to both empirical economic literature and the theoretical literature of evaluation research.

As for "costs," not every significant effect of the projects could be measured, due to institutional and methodological problems. (Areas of insufficient knowledge include health impacts, city-wide effects, and the significance of income and employment generation impacts). Delays in evaluation data processing held up the investigation of important theoretical questions. A full benefit-cost accounting of every component of the evaluation has not been attempted. The benefits cited, however, along with clarification of project concepts, and the demonstrations of evaluation's utility, make the DEDRB evaluation a significant contribution to an understanding of the complex urban sector in the developing world.

Chapter 10:

DESIGNING EVALUATIONS

The Nature and Appropriate Roles of Evaluation

World Bank senior management has justified the institution's extremely cautious introduction of systematic evaluation with the opinions that it is not clear what evaluation is (or should be), and that, whatever it is, there are very few persons qualified to do it.

If these assertions had some validity in the early 1970s, by the end of the decade - as the foregoing chapters clearly demonstrate - they had none. During the 1970s, evaluation became a growth industry in North America, with many competing practitioners. Furthermore, joining their ranks has required the mastery of no arcane science or techniques. What has been lacking in the field has been good judgement in adopting suitable blends of objectives and techniques.

So, evaluation and evaluators certainly exist. The bulk of this book demonstrates that the Bank's evaluation program in the urban sector has been piloted successfully through the rushing straits of burgeoning evaluation activity in the 1970s. It has been caught up in a few eddies, and has had to toss some cherished objectives overboard. But it has ultimately sailed calmly out with very good marks in all major fields of evaluation performance. It scores very high in terms of maintaining and strengthening modeling and implementation paradigms to frame the research. This fact has been principally due to early recognition that studies of goal attainment in themselves would fail to capture reasons why social and economic programs have the effects they

do. Thus the program has out-performed most others in tracking project implementation and explaining what has happened.

If there was a significant overall weakness in the Bank's "urban evaluation model" it was in the area of clarifying evaluation uses and specifying its users. In the early going, concentration on getting the entire apparatus in place caused neglect of some of the immediate needs of project managers, and even now much can be done to improve the formats of management reporting and other dissemination procedures. Nevertheless, here too the program scores well, relative to most others.

Those who believe that simpler and accurate evaluation methods are at hand should read with care the considerable literature on evaluability assessment, rapid feedback evaluation, and performance monitoring. These methods are useful, notably for hypothesis generation, reformulation and refinement. But, although we have used them and recommend their continued use they cannot stand on their own.

There is no ambiguity concerning the DEDRB program's performance in the third major area - refinement of measurements and indicators. Here, as in the first category, the record has been very strong, and that of comparable programs very much less so.

Finally this book - along with the complementary set of seven manuals simultaneously prepared by my colleague Michael Bamberger - testifies to the arrival of meta-evaluation, and provides the basis for making some recommendations about the use and roles of evaluation in the future. This will be done in the following sections of this final

chapter: the first recapitulating the "case" for evaluation; the second clarifying the bureaucratic milieu into which it must be introduced; the third discussing the establishment of objectives; the fourth laying out the principal components of the recommended evaluation system; and the last supplying guidelines for the organization and execution of the program.

Gains from Evaluation: The "Case" Reaffirmed

The main justification for evaluation of social programs is that it enhances the learning processes they involve, which is vital in a rapidly changing world. Evaluation does this principally by making the process more certain, thus increasing everyone's confidence in the actions being advocated, and by speeding up the processes by which learning takes place, after a certain "running in" period. The first facet is perhaps the more obvious, and there is little doubt even among critics that systematic evaluation can increase confidence in findings, the more so as research follows a rigorous model. It is worth stating the reason why this is so important. Systematic evaluation - and the increased confidence it makes possible - takes on more importance the further removed is the person who must act on findings from first-hand familiarity with the project supplying them. In short, rigor helps findings stand the tests of time, distance, and changes in cultural, institutional, and policy contexts.

Those who doubt that evaluation will actually speed the accretion of knowledge, after an initial period, have simply failed to examine the record, or else they fail to understand the investment

process (the investment here being in a more efficient learning model), which always involves an early waiting stage.

Evaluation can also make important contributions to conceptualization and clear thinking about programs and processes; that is, about what projects can and cannot achieve.^{1/} This kind of contribution is particularly significant in a case such as that of rural development initiatives aimed at alleviating poverty, where, in addition to resource constraints, there are conflicts among objectives, uncertainty about technological packages, weak administrative structures, and incomplete knowledge about the actual effects of different policies on the people they are meant to benefit. In the urban sector, the World Bank's evaluation program has greatly aided conceptualization and understanding in the areas of self-help, income distribution and the role of transfers, costs and credit requirements, and other issues.

With better understanding of such issues comes greater consistency and continuity in investment programming. For example, USAID entered the low-income housing field several years before the World Bank did. It did so misunderstanding many things with respect to ends and means. When its first project, introduced in Panama in 1964, failed to bring about the desired results, the program was discontinued in that country, as part of a general cutback of such USAID programs

^{1/} The negative specification arising from evaluation may be even more important sometimes than its constructive aids, for it can identify conditions which cannot be altered or improved, and thus help policy makers avoid waiting time in a search for solutions that don't exist.

toward the end of that decade. This legacy of hasty introduction and even hastier withdrawal made it very difficult for USAID to introduce a new housing project in the latter 1970s, by which time it had been convinced that approaches other than those it had chosen had a much better chance of working.

In the Bank's evaluation programs and in the literature generally, evaluation is set three tasks: solving problems, measuring impacts, and creating memories. Though the respective weights of these three aims have seldom, if ever, been clearly defined, the Bank's program makes it clear that evaluation can make important contributions in all three areas. These contributions take the form of clarifications of thinking, "tracking" ^{1/} of project progress and the improvement of measurements (e.g., for cost-benefit calculations) and record keeping. With respect to the last two factors, evaluation can enable us to get more out of existing data, not only by adding critical complementary items, but also simply by effectively documenting that which is being generated anyway.

The General Research Advisory Panel (GRAP) which assessed the World Bank's research program during (), stated that "although the inadequacy of data is often cited as a major constraint on empirical research, it is also the case that the analytical potential of such data as exist is not fully exploited." ^{2/} This has been true for a

^{1/} "Tracking" is here defined as: (a) determining what is going on, (b) assessing reasons for deviations from the prescribed course, and (c) taking corrective measures.

^{2/} GRAP Report, p. 11.

combination of reasons, including the failures to spell out data requirements precisely and to document existing data effectively, so that it can be judiciously supplemented and efficiently used. In the development field, the failure to document and store for the future use of others, data collected at great expense approaches scandalous proportions. Since such documentation is hardly ever done, let alone done well, it is little wonder that few potential researchers or evaluators know what data are already available. It should be a high priority of all research and evaluation programs to improve the record in this regard.

By assuming such a role in data specification and documentation processes, appropriately designed evaluation systems can become the "missing link" between operations and research in institutions such as the World Bank, and thus contribute to the long-sought operationally oriented" research program. They can do this by providing ongoing, systematically generated and analyzed agenda of questions, with criteria for their ranking in terms of urgency and researchability. By making the provision of questions a continuous process, the evaluation effort can explain changes (e.g., of emphasis, composition, or direction) as they occur, adjusting the weighting of issues and the resources of the research accordingly. Such practice would enhance the relevance of research and help it improve tracking and

the basis for record keeping.^{1/} This approach would also aid the processes of project identification and appraisal by throwing up a fuller and more suitable range of options for projects and their components, and by providing fuller quantification for benefit-cost estimation.

Evaluation is useful also in that it trains managers and operational staff in optimal information gathering and use, in the identification and analysis of problems and opportunities, and in the explanation and articulation of findings and positions. Thus, evaluation helps provide more valid responses to criticisms, both in terms of acknowledging mistakes and areas of ignorance, and in terms of refuting false arguments, and claiming successes accurately.^{2/}

The above factors add up to a strong case for evaluation as being in the interest even of the "evaluated," the responsible policy makers and practitioners. However, even if this is not their perception (and the following section discusses some reasons why it often is not), there is a moral argument for evaluation: that it is the essence of professional responsibility to try to measure more accurately the consequences of our actions, even if we may never approach very close to an ideal. This argument applies particularly to development finance

^{1/} It would remain only for operational personnel to design and employ the associated recording instruments and process.

^{2/} It has been found that, both in the World Bank and the participating countries, evaluation has helped make the case for the kinds of urban lending strategies being advanced. This potential contribution was in fact early recognized by urban project management within the Bank, and accounts in part for their sustained support for evaluation in the face of persistent resistance by lower-level staff, and by managers in some other sectors within the Bank.

institutions, which advocate policy and program innovations in poor countries, and frequently do so in promising "new" areas of activity where ignorance abounds.

A Hostile Institutional Environment

The uneasy chemistry between evaluators and those evaluated is one reason why more evaluation of social sector programs has not taken place. Cynical evaluators, or would-be practitioners, may say that evaluation is difficult to prosecute because managers don't want anyone to know what is really happening, and therefore place many obstacles in the way of the effort. We did not find much evidence of this within the World Bank or in the countries involved in Bank-financed projects. Rather, the tendency might better be described as one in which managers do like to control information flow in order to insure that the "proper interpretation" is placed on what they are doing. This tendency appears to intensify as one goes downward in an organization, because responsible officials have a less diversified portfolio as one proceeds in this direction. This is so because perceived mistakes are the more damaging, the fewer generally recognized successes there are to cite on the other side of the balance.

A counterpart extreme view to the cynical one taken by some evaluators is that evaluation is superfluous. Operational personnel often aver that evaluation makes no sense because they have been learning for years without it. This too is a wrongheaded view, simply because there is no end in sight to our ignorance, and anything that can help make inroads into it ought to be considered. Even though it is

true that learning occurs without systematic evaluation, this learning is unnecessarily slow and uncertain, and some of what is learned in unsystematic ways is wrong. An example from this experience would be the erroneous views long held, and seemingly confirmed by experience, about the differences in income distributions between sites and services and squatter upgrading projects.

The extreme views of evaluation just outlined do not suffice to explain why there has been so little evaluation of social sector programs to date. The World Bank's experience, for example, in early urban lending activity, when three of the first ten projects approved became "problem projects" requiring substantial readjustments, provided evidence that much can be learned quickly without evaluation, even though one of these projects (Senegal) is still uncompleted. The basic question is whether the cost of unsystematic learning by head-on, ad hoc confrontation of unforeseen problems is worth sustaining if means can be devised, and have been devised, to reduce it. A related point has to do with the actual and potential benefits of various approaches to evaluation. In this connection, it needs to be pointed out that the generally accepted interpretation of some of the findings of the World Bank's Operations Evaluation Department is essentially erroneous. Bank management and staff tend to think in Panglossian fashion that all is ok simply because everything "comes out alright in the end." However, "alright" is defined as producing comparable internal rates of return, ex-post as opposed to ex-ante. It doesn't take much investigation to demonstrate that this is not in fact good enough. In Bank projects in

most sectors the net losses from delays and cost overruns are very considerable in an absolute sense. And little comfort should be drawn (though much tends to be) from the fact that these losses are offset by "unexpectedly" high benefits. Whether these benefit differences are indeed unexpected, or what is being recorded here is the result of very conservative appraisal, it should be possible -- with evaluation -- to correctly anticipate these higher benefits and cost, while minimizing at the same time truly unexpected costs. More important, it is virtually certain that the choice of projects to be implemented would be influenced by truer estimates of costs and benefits beforehand.

A useful view of the problem was advanced by a knowledgeable senior staff member during a recent review of monitoring and evaluation experience within the Bank. This individual viewed the institutional problem as being due less to open resistance than to a lack of understanding and a lack of time to apply evaluation. In this view, managers need both a clear presentation of how evaluation will help them solve their problems, and of how it can be applied without costing too much of their time. The first part of this statement is unexceptionable and reiterates the fact that evaluators need to give more attention to devising dissemination methods directed to the needs of managers. The second portion of the statement reveals a bias common to operational personnel and exhibited as sharply in the World Bank as in any other institution. Operational personnel have an extremely steep time preference curve. They typically insist on simple systems without clearly thinking through what they want these simple systems to yield

them, or what their relative cost efficiency might be compared to alternative methods that may be more difficult to apply but would produce much richer results. Thus, those who would evolve evaluation systems must attempt to understand the biases against evaluation the better to be able to separate fiction from fact.

It is essential to observe that there is an inherent conflict between what might be called the experimental and bureaucratic orders.^{1/} The experimental order, to which evaluation belongs, has aspects at variance with the logic of the bureaucratic order, which underlies the functioning of any organization. The notion of evaluation is meaningless unless the staff of the organization involved in the relevant research considers the strategy to be evaluated as experimental, rather than as a permanent or definitive solution to the problem at hand. While evaluation implies that solutions adopted are only approximations, and leaves room for possible errors, the bureaucratic vocabulary translates the term "error" into the concept of mistakes which can imply a judgement on the staff responsible for execution of a project.

But this "opposition of orders," though present, is not at the heart of the matter. The crucial issue appears to be the difference in inherent time discount rates. Operational staff, particularly in Robert McNamara's World Bank, have been akin to peasant farmers (the archetypal risk averters) whom experience has taught that no amount of increase in

^{1/} I am indebted to Remi Clignet for a much more complete statement of this issue than appears below.

current consumption (loan processing) in period two can compensate for any significant reduction during period one. In short, they tend to say: "If you can't tell me now, it won't help me to know." Evaluators, on the other hand, have quite a different view of the world, as their whole business is to invest in the attainment and refinement of practical knowledge. Then there is the matter of the limited attention managers and operational personnel are willing to give to research findings. Evaluators who neglect this factor in planning the essence and presentation of their programs will be doing so at their peril.

The evaluator who would succeed in the bureaucratic context must be prepared to combat two characteristic misconceptions: the view of researchers that operations personnel are always trying to get something for nothing, and the view of operations personnel that researchers always take far too much time and resources to produce anything. The first is within the evaluator's control. He or she has simply to remember always to try to meet operations personnel more than half way. The second demands effective interaction. Because the research being advocated is consuming of time and resources, risky, and long-term, the would-be evaluator has to find ways to "co-opt" skeptical operations personnel, perhaps by producing some operationally useful results at an early stage, by presenting both objectives and results in ways that can be easily grasped, etc.

It will be useful to mention a few major institutional idiosyncrasies of the World Bank for what they may suggest about the

definition given to evaluation within its context or about the tactics would-be evaluators may have to use to win acceptance for a program.

The World Bank has long been an institution whose principal raison d'être is to grant loans and disburse funds. It is not unique in this function, but may be pre-eminent. The Bank has tended to convince itself that it doesn't make mistakes, so the opposition of bureaucratic and experimental orders within it has tended to work against both evaluation and pilot projects, against the latter because they are too small to move enough dollars, and because they require a disproportionate amount of supervision per dollar lent. This is significant because absence of pilot projects removes one link in the optimal learning chain to which evaluation might be attached.^{1/}

Another idiosyncrasy has been the denial, under Robert McNamara, of any conflict between quality and quantity in the Bank's operations. This is definitely a peculiarity of World Bank practice. Its main effect on staff has been one of demoralization. It is also antithetical to evaluation, increasing the bias which is present in any organization toward quick and dirty solutions to problems, quick answers, etc.

Fortunately in the urban sector there were some countervailing forces working against the prevailing biases and leading to the adoption of a substantial experimental evaluation program. Probably most important among these was the fact that management within the urban sector needed help making their case for a lending program at

^{1/} See Suchman and Alkin (Ref. ?)

a time when rural development lending appeared to have the advantage. Another favorable factor was the attraction of the relatively new urban field for an uncharacteristically bright and open-minded group of staff. These staff realized their own ignorance and reacted to it by maintaining a skeptical attitude toward their own efforts and a reasonably open mind concerning evaluation. A negative feature was that, in a new and complex field, the much heavier than average supervision requirements competed with evaluation objectives in their demands on staff time.

Setting Evaluation Objectives

The first condition for designing effective evaluation programs is to recognize that not all projects can be evaluated in detail. Beyond a basic evaluation component built into a project's management information system,^{1/} only certain projects (no more than one in ten in most instances) should have the kind of elaborate evaluation system on which the recommendations of this book are based.^{2/} The basic system should be relatively simple, although it must retain its basic rigor of method, including the extended time period required to make many evaluation objectives credible. Even the simple, basic evaluation should be framed by the quasi-experimental design paradigm. With reference to the three principal objectives of evaluation, (1) problem solving, (2) measuring impacts, and (3) creating memories, World Bank management might be said to put a premium on (2), to hope for (1), and

^{1/} (Ref. to Mike's Manuals)

^{2/} As designers of the pilot evaluation program, we fully expected that would be the nature of our recommendations.

hardly to consider (3). Operational staff tend to be almost totally concerned with (1), and evaluation staff are most interested in (2). In fact, all three objectives are important, especially when put in the following time progression: starting with (1), to establish credibility, then proceeding with (3). Then (1) and (3) together form the basis for identifying problems, articulating hypotheses, and attacking (2) efficiently.

Care must be taken to see that an evaluation's reach does not exceed its grasp. This can be achieved by being clear about what evaluation can't do and shouldn't try to do. Then, a discriminating model must be applied even to what evaluation can do, because there are clear trade-offs between quick, selective results and a more comprehensive picture obtained over a greater span of time.

What is needed, then, is a pragmatic approach to problem solving, recording, and measurement, based on key variables and interrelationships carefully chosen beforehand. However, the system must also remain flexible, so that the evaluation will not slavishly follow variables and relationships once thought to be important, but later demonstrated by evaluation or observation no longer to be central. Data and data processing requirements should be tailored to the individual country situations. Constraints in this area will be relaxed considerably by the development of a broad range of microcomputers and associated software. Program design should incorporate appropriate ratios of supervision time to field staff time,

something that has not always been a feature of World Bank-supported monitoring and evaluation schemes.

Such a system would permit of a limited number of occasions for more detailed statistical analysis and associated research. Such research opportunities should be directed to key questions requiring improved quantification, and priority should be given to using existing data and data already being collected. To make this possible, the basic evaluation design, in addition to being rigorous, should be extensive enough in its baseline data accumulation to support a program of associated research on impact evaluation, should program and evaluation findings so dictate.

When we compare the foregoing recommendations stemming from the pilot evaluation program to the monitoring and evaluation concepts that have grown up in the Bank more generally, there are some obvious differences. The recommended program is much the less "goal attainment" and ex-post, benefit-cost oriented, concentrating instead on problem solving, learning, and recording results. Secondly, our definition of evaluation embraces monitoring, which we see as distinct from evaluation in the following sense: while evaluation denotes the process of analyzing various kinds of information with various aims in mind, monitoring is simply one category of information gathering. Specifically monitoring refers to gathering information on the same variable (for example, a patient's temperature, or total disbursements) at prescribed intervals.

While it may be obvious that our recommendations imply revamping and radical improvement of conventional project reporting, it is worth stating explicitly, since the point is an important one. The Bank's present reporting system (supervision reports and project completion reports) is far too goal-oriented, seemingly aimed at the capacity to claim whenever possible that projects' original goals have been attained, regardless of the path taken to them, and consequently revealing next to nothing about what has actually happened. This approach is unsatisfactory for any development institution, both because delays, cost overruns, and benefit "misestimates" matter greatly, and because the process of "learning by doing" is seriously compromised.

A final point about selectivity is that one should probably not start in most instances to evaluate with the first project implemented in a given country, because the program slippage is likely to be considerable and evaluation resources, which have to be lined up beforehand, will tend to be wasted. With first projects, the evaluation effort should be limited to very basic record-keeping, leading to generation, reformulation and refinement of hypotheses. An early start should be made, however, in planning an evaluation for the second project, to commence at its outset.^{1/}

^{1/} Our experience offered two exceptions: the El Salvador exception derived from the fact that the executing agency already had a very effective management information system and a kind of basic evaluation program before the Bank entered the picture as funding agency. The Philippines exception is based on a pre-existing exceptionally good relationship between the evaluation program manager and the managers of the National Housing Authority in the Philippines, coupled with what is perhaps the most favorable position in practically any developing country with respect to the availability and cost of evaluation.

The evaluation objectives outlined here can be achieved in most countries at the expenditure of from \$50,000 to \$75,000 per (large) project per year. This is somewhat less than the actual costs incurred in the pilot program, even allowing for inflation, because of the simpler methodology advocated. It is a much less costly model than is currently being applied in programs defined as monitoring and evaluation in the development assistance field, and still less costly when compared to the major human services evaluation experiences in the United States during the last decades. Most of those programs each have cost several million dollars.

The Role of Evaluation: Planning and Evaluation Program

Evaluation programs should enable officials responsible for investment projects and programs to record the course of events and, in so far as possible, their causes. The activities involved should contribute both to more accurate "tracking" and "steering" of current programs and to the design of future ones. Both the analysis of problems and the selection of better projects (with a better mix of components) involve some quantification. In addressing these objectives, the evaluation re-defines and enriches the management information system. By a slight additional (and optional) step it can also improve considerably the base for associated research.

Nearly a decade's experience in planning, implementing, and assessing monitoring and evaluation of urban shelter programs has led to a fairly comprehensive, though tentative, set of views concerning the design and conduct of evaluation systems. These views naturally relate

Table 10-1: THE EVALUATION SERIES

Volume	Title	Contents	Audience
1	Planning an Evaluation System for an Urban Shelter Program: Key Issues for Program Managers	Definition of ^{the} main users of M&E and the types of information they need. ^{Key issues} in each type of study ^{and with major} research designs and main stages in the planning of an evaluation system.	Policy makers and program managers
2	Methodology for the evaluation of urban shelter programs -- an overview.	Review of research techniques for the design, implementation and interpretation of evaluation systems.	Researchers
3	A basic evaluation package	Presentation of a basic evaluation system with a step by step explanation of how it should be set up.	Program managers and researchers
4	Designing a questionnaire for longitudinal impact studies	Discussion of the main types of information required for measuring project impact over time. Examples of typical questionnaire formats.	Researchers
5	Non-survey techniques in the evaluation of urban shelter programs	Description of non-survey ^{evaluation} techniques which avoid many of the problems inherent in the use of structured questionnaires.	Researchers
6	Statistical evaluation of project impact through longitudinal surveys	Review of the main statistical techniques for evaluating project impact with different types of sampling design. Explanations are included on how to use each technique.	Researchers
7	Computer Analysis of Longitudinal Impact Studies: Some Issues	Discussion of typical problems with matching cases, data cleaning, consistency checks and data transference from one computer	Researchers

to and draw their examples from shelter schemes oriented toward low-income populations. However, we have yet to discover any reason for believing that they are not applicable to other sectors. In addition to the narrative and descriptive analysis presented above of evaluations carried out in four countries, the current output of the program includes seven detailed modules designed for different user audiences among those concerned with urban shelter projects. This section presents an overview, based on our experience, of how these modules should be used (See Table 10-1 for a list of the modules).

Aims The specific steps that should be taken to plan an evaluation are outlined in Figure 10-1. As noted in the first step, initial management decisions ought to involve a firm assessment of the kinds of information needed throughout the project process. This stage should also include recruitment of the primary evaluation researchers, establishment of the organizational linkages between management and evaluators, development of a plan for the scheduling of evaluation efforts, and definition of budget and staff requirements.^{1/} With the naming of the evaluation team, the work plan specifying the translation of information needs into a research design must be undertaken. This

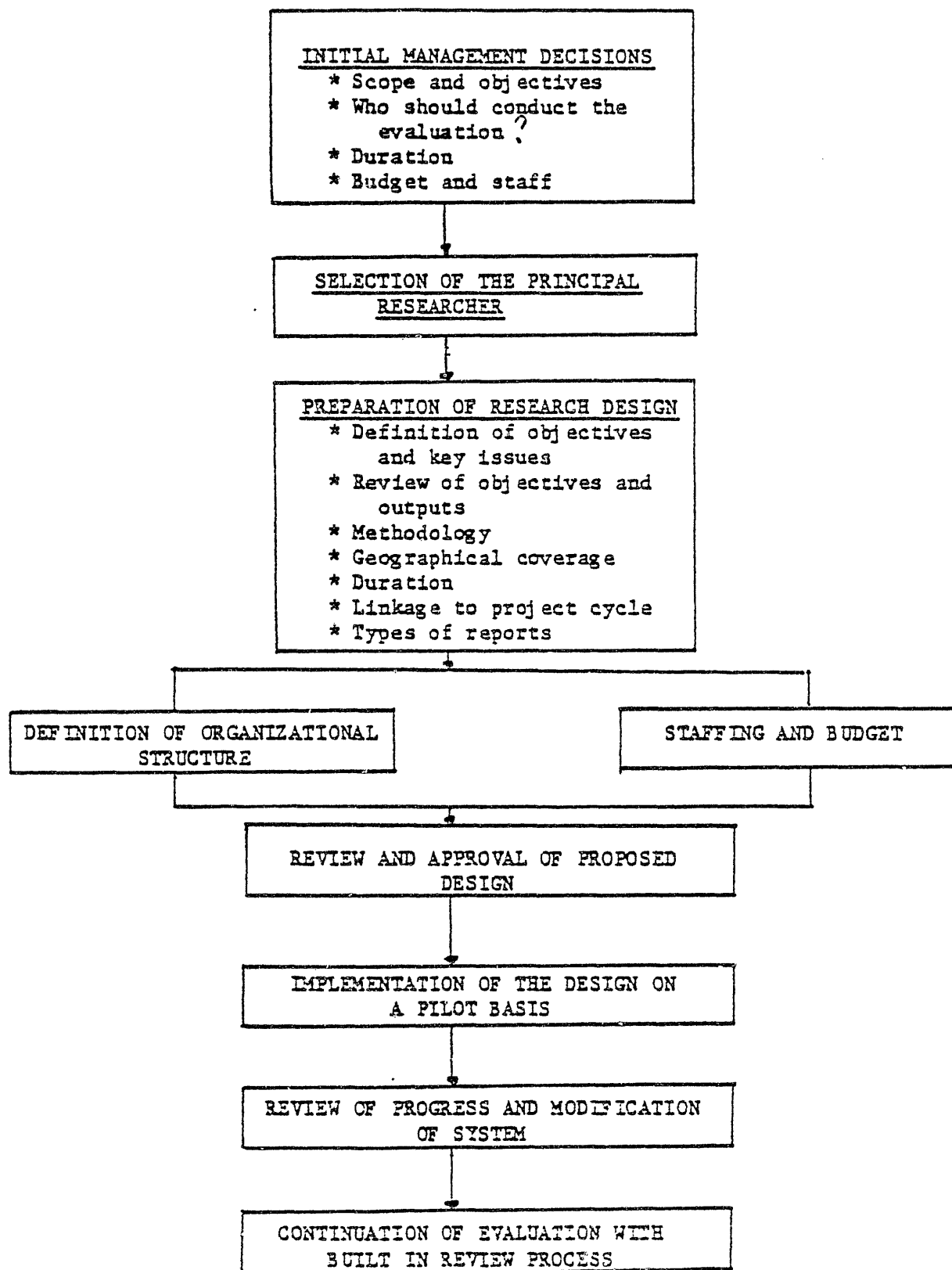
^{1/} Our experience has shown that employing host country nationals rather than expatriates as researchers contributes to developing lasting commitments to the evaluation early on, and facilitates drawing on accumulated professional and organizational expertise within the country. In most cases, two to three years appear to be required to develop a research team capable of meeting the relevant range of management information needs. Consultants are best used not to manage, but to train or enhance nationals' expertise, so that minimal capability is sacrificed if an engaged consultant leaves the project.

task includes definition and review of the evaluation's objectives and key issues, as well as the kinds of methods, geographical coverage, timing and reporting procedures to be used. Managers must be prepared to say how finely tuned they desire any particular study to be, or run the risk of misallocating evaluators' time in too detailed or circumspect an effort. Review of the organizational and informational links between evaluators and managers should then be initiated again, to make sure that adequate staff and budgets have been allocated for the desired tasks. As a continuation of this iterative design process, the proposed research plan should be examined and approved by relevant audiences; tested on a six-month or other pilot basis, to see that expectations for the design can be fulfilled; reviewed and modified if gaps or new needs have been identified; and, as amended, continued with a periodic, built-in review of the evaluation's adequacy.^{1/}

During the articulation of the evaluation design, the evaluation itself will identify needs for specific studies aimed at project management and other audiences. The points where this occurs are graphed along a project planning--implementation--new project planning continuum in Figure 10-2. At the planning and design stages, managers need most to know about the characteristics of housing demand

^{1/} Failure to involve program management in the evaluation's design exacerbates the misperception of evaluation as an "expert's" domain, and risks diminishing the evaluation's utility to project personnel. If implementers and managers understand and agree upon the bases on which data will be collected (on specific goals, issues, progress toward objectives, etc.), less time is likely to be lost, should an unforeseen problem require adjustments or new steps in using evaluative data to solve it.

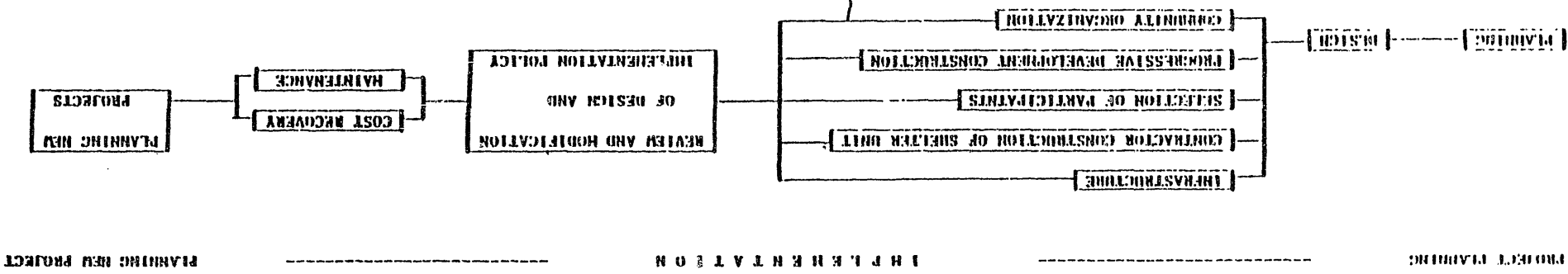
Figure 10-1: MAIN STAGES IN PLANNING AN EVALUATION SYSTEM



in particular cities or neighborhoods; the affordability of projected housing options for low-income populations; the range of housing and associated services they require; tenure options; optimal locations for project sites; employment generation possibilities of the projects; and design requirements, such as costs of materials, construction methods, service layouts, and credit demand. During implementation, managers need to keep abreast of the progress of project-wide infrastructure; contractor construction of housing units; participant selection procedures; advances in families' own progressive development of their housing; and community organization.

Each of these components, must be gauged against planned levels of physical progress, costs, and disbursements, with time taken to relieve bottlenecks and solve problems. Assessment must also be made of the efficiency of implementation and the quality of houses and services. If these are found wanting, modifications ought to be made in planning and design features to insure continued quality control. Before the end of project implementation, evaluation must turn to the procedures for cost recovery and maintenance of project features, and to initial impacts on participating low-income families, on the cities involved, and on national housing policies. At appropriate intervals during the overlapping processes of new project design, there should be a "stock-taking" of accumulated experience and evaluation findings.

Figure 10-2: INFORMATION REQUIREMENTS AT EACH STAGE OF PROJECT DEVELOPMENT



10-24

Type of Study	Planning	Project	Quality Control	Accountability	Control of Progress	Control of Program	Control of Budget	Control of Construction	Control of Operation
Demand	Impact on participants	Impact on low-income families	Efficiency of implementation	Comparison with objectives and implementation	Physical program	Physical program	Physical program	Physical program	Physical program
Affordability	Impact on housing policies	Impact on the city	Quality of houses and services	Comparison with objectives and implementation	Physical program	Physical program	Physical program	Physical program	Physical program
Types of services	Impact on housing policies	Impact on the city	Quality of houses and services	Comparison with objectives and implementation	Physical program	Physical program	Physical program	Physical program	Physical program
Facilities	Impact on housing policies	Impact on the city	Quality of houses and services	Comparison with objectives and implementation	Physical program	Physical program	Physical program	Physical program	Physical program
Employment	Impact on housing policies	Impact on the city	Quality of houses and services	Comparison with objectives and implementation	Physical program	Physical program	Physical program	Physical program	Physical program
Components	Impact on housing policies	Impact on the city	Quality of houses and services	Comparison with objectives and implementation	Physical program	Physical program	Physical program	Physical program	Physical program

All the studies associated with Figure 10-2 are not required for every project, nor by every project manager. Different projects will have different aims, and flexibility is necessary to tailor evaluation assistance to assure that the needs of specific country housing strategies are met. The studies and iterative evaluation design together represent a comprehensive attempt to meet the data and organizational needs of project managers, and so to maximize the usefulness of evaluation research.

The same information requirements for the conduct of urban shelter projects can be organized in a slightly different manner, as shown in Table 10-2. Here, the concern is: (1) to chart project progress, or the conversion of project inputs into outputs; (2) to gauge the effectiveness of the outputs; (3) to measure input/output ratios over anticipated cost and time, and compare them to those for alternative shelter designs (that is, to assess "internal" and "external" project efficiency); and (4) to relate project outcomes to general planning for other urban projects. These lists of "indicators," all of which have been discussed in previous chapters, should not be taken as constituting a necessary or sufficient catalog of the variables involved in low-cost urban shelter programming for the urban poor. However, they are all important features of any evaluation attempting to gauge the efficiency and effectiveness of housing schemes.

The four sets of indicators summarized in Table 10-2 are not intended to be directed exclusively toward project managers. Rather, they are intended for the full range of development practitioners

Table 10-2: THE MAIN TYPES OF INFORMATION AND ISSUES WHICH
EVALUATION RESEARCH CAN COVER

1. INDICATORS OF PROJECT PROGRESS (INPUTS - OUTPUTS)

selection of project participants
construction
occupancy
house consolidation
drop-outs
maintenance
cost recovery

2. INDICATORS OF PROJECT EFFECTIVENESS

affordability
accessibility
increasing housing stock
improving access to urban services
impacts on target populations
effect on urban housing markets
effects on development policies

3. PROJECT EFFICIENCY

Efficiency of Individual Project Components

project planning & design
selection procedures
construction methods
loans
maintenance
cost recovery
community participation
plot occupation

General Project Efficiency

efficiency in terms of project goals

design
finance
implementation
maintenance
cost recovery

comparison with alternative shelter programs

cost comparison
quality comparison
replicability

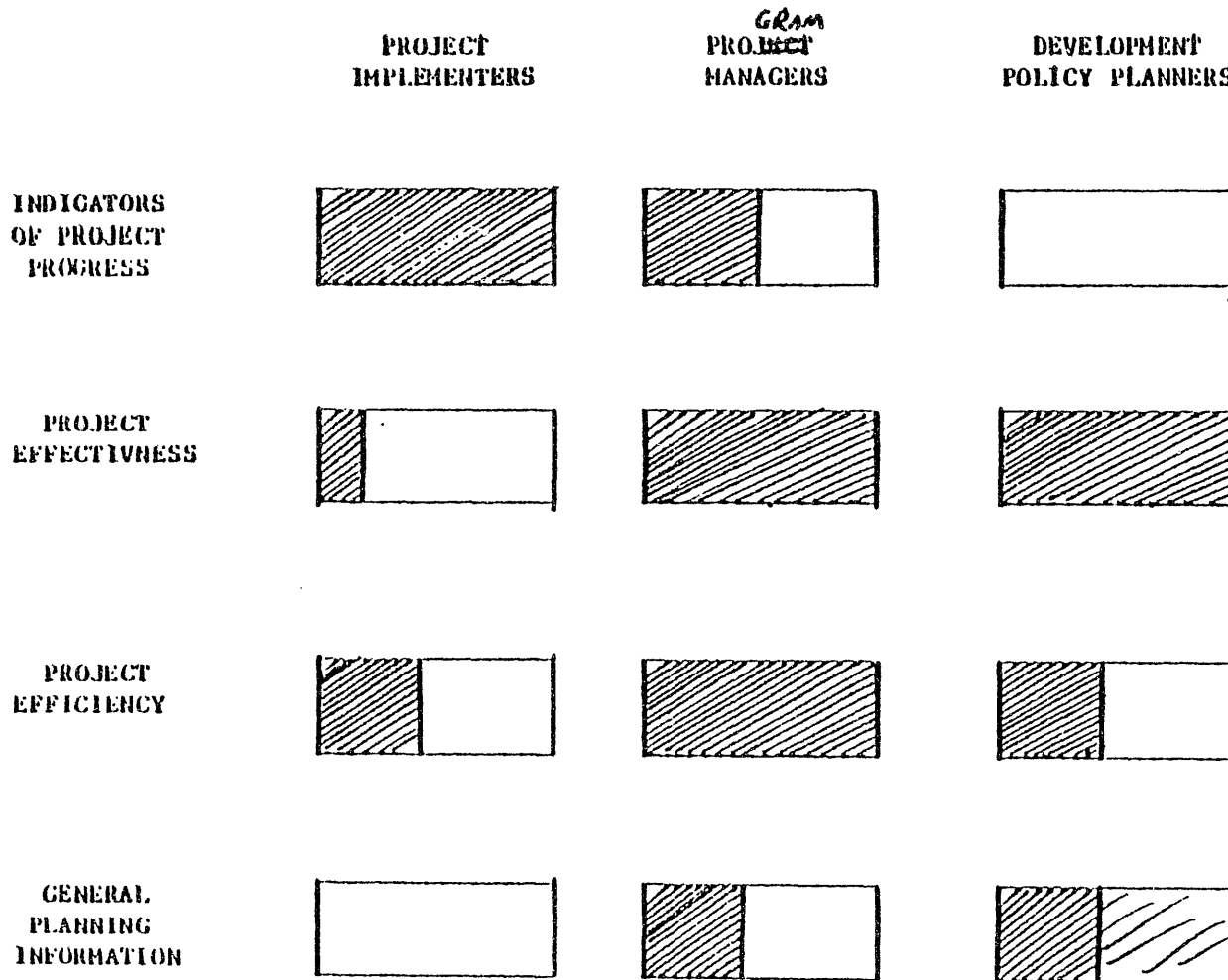
4. GENERAL PLANNING INFORMATION

income and expenditure
employment
demographic

involved in urban shelter schemes, especially for the project implementers, project and program managers, and higher-level policy makers or planners who will use the information in designing new interventions. Their relative interests in each of the four categories of information are illustrated in Figure 10-3. As noted, project implementers will be most concerned with indicators of project progress, to some extent interested in those measuring effectiveness, and to a somewhat greater extent in those gauging efficiency. They are not likely to be consumers of general planning information. Program managers will be significantly interested in progress indicators, but their primary consumption of information will be in the areas of project effectiveness and efficiency, as these domains are their principal responsibility. They also have a stake in the planning information, which would be used to formulate new efforts based on experiences with projects under their direction. Policy makers and project planners are not liable to be major audiences for project progress studies, but are more interested in project efficiency and are major consumers of reports detailing project effectiveness and the scope of planning information.

This overall framework is not devised to exclude any potential user from access to data on all indicators, but rather to help managers streamline information flows by channeling categories of information to those whose functions most require them. If information requirements are kept relatively simple, expanding only as problems dictate, evaluation can serve as a flexible management tool, buttressing normal reporting procedures.

Figure 10-3: THE PRIMARY CONSUMERS OF EVALUATION RESEARCH AND THE KINDS OF STUDY IN WHICH EACH IS MAINLY INTERESTED



KEY: Shaded proportion of rectangle indicates the degree of interest.

Methods

Having covered the aims, we shall now deal with the methods of evaluation. Procedures and resources, rather than questions of precise methodology are what are involved here. It is relatively easy to define the bounds of the methods to be used. All projects should have better reporting systems than they do today. Improvements should be based on process evaluation, involving "in course" corrections, with each step explained as it is taken. Methodological purists who say this approach biases the results of evaluation^{1/} completely miss the point. Such a bias might develop if the reasons for altering the course of a project were idiosyncratic, but as long as they are not, then the evaluation will not suffer. For example, when it was found that introducing a credit scheme into the Senegal project improved the response of some of the population, its initial absence could be recorded as a design error which has corrected so that the project could proceed. Thereafter, the evaluation could concentrate on the pros and cons of the specific credit scheme. Such adjustments ensure that all projects have basic evaluation and improved reporting at the project level.

At the program level (the Bank's evaluation program applied to a whole sector of activities across countries), a small selection of projects, or sub-projects, should have built into them some detailed examination of socio-economic issues or other areas of potentially

^{1/} (Ref)

consequential ignorance. These activities would involve combinations of techniques, including statistical and participant observer, and possibly longitudinal impact analyses, special studies, and associated research. Between the two levels, there is wide latitude for the exercise of evaluators' judgment, but not much basis in experience to date. The key, however, is that even the simplest evaluation must be framed in the quasi-experimental mode from the outset. This procedure gives order to all deductions, however crude the methods, small the samples, or statistically "insignificant" the results. In this approach, the early rounds of evaluation are concerned mainly with selection and refinement of hypotheses. Another important reason for having a solid methodological framework from the start is to lay down some baseline data for eventual use, even if no use is made of it for as much as five years.

A final observation on method is that one looks for a relatively fixed sample size for statistical validity, irrespective of the total population involved. Therefore, the larger the population to which findings apply, the more "efficient" the evaluation. Care should be taken to select large areas and/or populations for full-blown statistical evaluations, tied if possible to city-wide sample frames.

With respect to the conditions for evaluation, the main requirements are time, resources, and sufficient independence to maintain rigor and to avoid being forced to devote all resources to supervision or to implementation problems in the short term. The design for a particular country, sector, or project must be shaped by the

country conditions, including available data and data generating capacity; human resources in the social sciences; the efficiency and flexibility of involved institutions; and computer systems. This last factor has been a large impediment to most programs for developing countries to date. We expect a sea of change in this area, thanks to the growing availability of microcomputers and associated software.

Taking the long view is most important. It is wise not to commit resources before one is reasonably certain that implementation is in fact going forward, and thus, in general, to avoid assigning substantial evaluation resources to first projects. Yet, from the point of view of project planning, simple, basic evaluation should be introduced at the earliest stages even of first projects, for here preliminary surveys can improve project assumptions, provide insights into project design, and reduce overall evaluation costs.

The investigation should be focused with crude tools before moving to the use of more sophisticated analytical procedures. This is a way of avoiding "excess data," which was a problem with the DEDRB evaluation and with others we have reviewed. It is important to note that the "excess data" problem arises not because of excessive sample size, but because of a combination of too many variables, and possibly also too frequent repetition of survey instruments. If resources are committed to substantial evaluation only after principal problems have been agreed upon and implementation problems have largely been solved, one can proceed with much more confidence, and with the advantage of streamlined instruments, and accurate planning as to the

appropriate intervals for their application. Such efficiency will not necessarily save on data collection costs, since the basic issue here is to increase data reliability and efficiency of data use. However, it will lead to economies at all other stages of an evaluation exercise.

Evaluation should be understood not only as a way of producing certain kinds of results, but as a process of developing the capacity to answer vital questions. Such an effort involves creating organizations, developing staff and capabilities for data generation and processing, and developing management applications for findings, which will feed back to the beginning of the process.

Organization and Execution of Evaluation

This concluding section deals with the principles involved in organization for an evaluation in the field and in the lending institution. A first principle is that evaluation must be very close to management. The exact relationship will vary according to the organization involved, and according to whether one considers the issue at the project or the program level. (For example, program evaluation will be responsible to managers at a higher level than project management, and will thus inevitably have an "over the shoulder" quality from the project manager's viewpoint. The World Bank's involvement with evaluation is in some respects a case of this kind.) In placing evaluation close to program or project management, the need develops to provide some assurance of objectivity. Providing some such guarantee (such as oversight by the lending institution, oversight by a broad-based steering committee, continuous interaction between parties, etc.)

has proved to be far more effective than the alternative of assigning evaluation to an independent institution. A measure of this nature is also the only effective way to achieve the desired symbiotic relationships between managers and evaluators. Such an arrangement was achieved quite readily in El Salvador and the Philippines, and eventually in Zambia as well. That it was not achieved to the same degree in Senegal says more about the nature of that project's peculiar problems than about particular approaches to evaluation.

The need for conceptual and organization distinctions between monitoring and evaluation is obviated if methods for reporting evaluation results and all other information central to project decision making are clearly devised at the project planning stage. Evaluation experience in World Bank supported projects indicates that the designation of one team to monitor indicators of project progress and another to assess their findings leads to lack of coordination and duplication of effort. Hence, it makes it much more difficult to move efficiently between problem solving, impart education and design modification. Also, units with the sole responsibility for monitoring are unnecessarily exposed to implementation pressures and are the more easily drawn into ad hoc problem solving and project reporting roles, further complicating the difficulty of coordination with the principal evaluation activity.

Adequate planning is the key to budgeting staff time appropriately to meet project information needs in the short, middle, and long term. Unlike some others, we came to this conclusion very

early in the evaluation planning process. By now there is general agreement among evaluators and field managers that this is the way to proceed. Given the location of the evaluation unit in the responsible agency what will matter most is the competence of management and the evaluation team, the amount of interaction and communication between them, and a shared interest in learning.

Because there are many considerations which transcend project time periods (such as maintenance and long-term impacts), project scope and period and evaluation scope and period usually do not correspond exactly. The result is a need to find ways of making project finance of evaluation "stretch" differently. This seems not to have happened yet, but it should be possible as is done, for example, for technical assistance and studies within the World Bank's typical sequence of project lending activities in a country.

It is only appropriate to ask how an organization can satisfactorily evaluate itself. In the type of exercise we recommend, the organization will not really be evaluating itself, but only some of its activities. Still, there is no reason why an organization should be particularly objective about such an effort, unless it is under very severe pressure from the market, or from some other source, to improve its performance. In our experience, evaluations are best kept "honest" by judicious use of such devices as a national steering committee, Bank supervision, and peer review.

How should an institution like the World Bank define its own evaluation requirements and organize itself to meet them? An early

preliminary step would be for agencies such as the Bank to persuade national census and statistical organizations to collect necessary data so that city sample frames, as well as broad population characteristics, will be available when the time comes to plan research concretely. In this regard, current practice remains very far from ideal. It stems in part from the fact that the Bank does not have a policy on evaluation.

The fact that the present system in the Bank is very nearly chaotic stems from a fragmentation that can be traced to management failure. There has been a general underestimation of the task of evaluation and poor definition and selection of evaluation which are not. Certain other activities which are kinds of evaluation are not generally appreciated as such. My guess would be that very few people within the World Bank have an accurate picture of the way evaluation is occurring and of how inadequately it is coordinated. Most staff probably perceive that it is going on in the Operations Evaluation Department and within certain central sector departments. Fewer will be aware that such work is going on in the Development Economics Department. Probably almost none realize that it is going on to some extent throughout the institution. This situation leads to much duplication of effort, as well as to certain things falling between stools. Equally significant is the fact that the general lack of coordination contributes to a lack of coordination of data requirements. Here too duplications and omissions can both be very costly. All in all, the process is quite ineffective, because too few

people are working on it, because they are too scattered and because there is no "orchestration" of their work.

Not surprisingly, portions of the activities that are called evaluation are misdirected. The Operations Evaluation Department, for example, demonstrably wastes considerable time on incremental improvements to project completion reports, while much of what is done under the auspices of Central Projects Staff involves writing monitoring and evaluation requirements into projects, and then attempting to cope with a mountain of poorly collected data and poorly formulated reports from the field.

What is needed is an evaluation program that is defined for the institution as a whole, spelled out clearly and simply, placed under central direction, and staffed with adequate resources, consisting mainly of "doers," rather than advisers. Central direction makes possible the establishment of a framework and general methodology to be used throughout the institution. It also facilitates the setting of priorities, by problem or sector, or by level or type of analysis. Central direction is also essential for determining anything like optimal data requirements, and for appropriate assignment of responsibility for data collection and documentation, so that at later stages collectors and users can exchange data as necessary.

A central operation makes it possible to establish reporting requirements and to see that they relate appropriately to the goals of field work supervision and feedback of useful lessons into the institution as a whole. Other advantages of central direction include

creating a critical mass (both analytically and bureaucratically), providing the flexibility to change priorities and the balance of work between sectors and between types of problems, and facilitating inquiry into long-term issues that transcend project life.

If evaluators are doers, rather than advisers, mistakes tend to be detected early, and others within the institution will be spared the error of thinking that evaluation is an advisory exercise which can be performed well with only a few people working on it. The problem here is simply the steep time preference of staff and managers. This issue leads some to believe that evaluation must be imposed to be adequate, at least initially.

How effective evaluation is depends finally upon management's definition of the institution's key goals. In the World Bank's case, if the key goal is to push out the loans, would-be evaluators will continue to have a tough row to hoe. But if a higher premium is put on the speed of the learning process, and more emphasis is given to technical assistance as the institution's main product, then there is an easier road ahead for systematic evaluation.

With respect to agencies directly dependent on government funding, such as USAID, relations with federal institutions may determine the future of evaluation's growth. If the Congress is viewed as the enemy good evaluation will be driven out by public relations (as indeed is the tendency sometimes within the World Bank). On the other hand, if Congress is treated as an important part of a productive team, things become essentially more workable.

Perhaps most important is the fact that the subjects of the current leading research questions are remarkably consistent with the changing orientations of the World Bank's urban lending program, and with the interim conclusions of the evaluation work. This is more evidence, of the most convincing sort, of how evaluation serves concretely the needs of national housing authorities and development institutions such as the World Bank in their cooperative search for improved solutions to the housing crises of the developing world.