

AGENCY FOR INTERNATIONAL DEVELOPMENT WASHINGTON, D. C. 20523 BIBLIOGRAPHIC INPUT SHEET		FOR AID USE ONLY <i>Batch 72</i>
1. SUBJECT CLASSIFICATION	A. PRIMARY Development and economics	DC00-0000-0000
	B. SECONDARY Development assistance	
2. TITLE AND SUBTITLE Readings in project design		
3. AUTHOR(S) Stout, Russell		
4. DOCUMENT DATE 1977	5. NUMBER OF PAGES 78p.	6. ARC NUMBER ARC
7. REFERENCE ORGANIZATION NAME AND ADDRESS MUCIA		
8. SUPPLEMENTARY NOTES (<i>Sponsoring Organization, Publishers, Availability</i>)		
9. ABSTRACT This collection of papers constitutes a module for the study of project design, the study of methods and difficulties of program evaluation, and the application of a logical framework to program evaluation. "Foreign Aid: A Critique" provides an overview of the problems of aid administration and examines the link between project purpose and overall programmatic goals. The World Bank's programming process is examined in "The Project Cycle;" observations made in it are applicable to other institutions. The "Problems of Development Management" are assessed. It notes an increasing need for flexible, adaptive management, and points out that there is a tendency toward centralized decision making. The problem of the political contamination of project goals is also addressed. "Program Management and the Federal Evaluator" focuses on chronic difficulties in evaluating social programs and suggests that these programs often suffer from lack of definition, lack of clear logic, and lack of management. In fact some programs may not be evaluable and this must be communicated to policy makers and program managers. The same problems are addressed in "Formulating the Question and Measuring the Answer" where it is emphasized that lack of clear definition of program goals is a key problem in evaluating program success. "What Can We Actually Get From Program Evaluation" examines the potential value of evaluation of national programs and reviews the difficulty of meaningful evaluations of specific federal programs. "Principles and Methods of Program Evaluation" suggests a logical framework for program evaluation.		
10. CONTROL NUMBER <i>PN-AAE-988</i>		11. PRICE OF DOCUMENT
12. DESCRIPTORS Reference books Project planning Project management Project evaluation Foreign aid Technical assistance		13. PROJECT NUMBER
		14. CONTRACT NUMBER AID/otr-C-1421 GTS
		15. TYPE OF DOCUMENT

READINGS IN PROJECT DESIGN

Edited by
Russell Stout, Jr.

This material has been produced cooperatively by the
Professional Studies and Career Development Division,
United States Agency for International Development and
PASITAM, the Program of Advanced Studies in Institution Building
and Technical Assistance Methodology of
MUCIA, the Midwest Universities Consortium for International Activities



PASITAM: c/o Geology 545, Indiana University, Bloomington, Indiana 47401 U.S.A.

Grateful acknowledgment is made for permission to use the following:

"Foreign Aid: A Critique," by Albert O. Hirschman and Richard M. Bird. Reprinted from Albert O. Hirschman and Richard M. Bird, *Foreign Aid: A Critique and a Proposal*. Essays in International Finance No. 69, July 1968. Copyright (c) 1968. Reprinted by permission of the International Finance Section of Princeton University.

"The Project Cycle," by Warren C. Baum. Reprinted from *Finance and Development*, June 1970, pp. 2-13.

"Project Planning and Implementation in Developing Countries: An Introduction," by Dennis A. Rondinelli. Reprinted from *Project Planning and Implementation in Developing Countries* by Dennis A. Rondinelli and Aspy Palla by permission of the East-West Technology and Development Institute. Copyright (c) 1976, East-West Technology and Development Institute.

"Program Management and the Federal Evaluator," by Pamela Horst, Joe N. Nay, John W. Scanlon, and Joseph S. Wholey. Reprinted from *Public Administration Review*, vol. 34, no. 4, July/August 1974 by permission of the *Public Administration Review* and the authors. Copyright (c) 1974, Public Administration Review

"Formulating the Question and Measuring the Answer," by Carol H. Weiss. From Carol H. Weiss, *Evaluation Research. Methods for Assessing Program Effectiveness*, (c) 1972, pp. 24-59. Reprinted by permission of Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

"What Can We Actually Get from Program Evaluation?" by Joseph S. Wholey. Reprinted from *Policy Sciences*, vol. 3, September 1972, pp. 361-369 by permission of Elsevier Scientific Publishing Company. Copyright (c) 1972, Elsevier Scientific Publishing Company.

"Principles and Methods of Program Evaluation," by Herbert D. Turner. Reprinted from *International Development Review*, Focus section, 1976/3 by permission of the Society for International Development. Copyright (c) 1976, Society for International Development.

Contents

Project Design and the Logical Framework <i>Russell Stout, Jr.</i>	1
Foreign Aid: A Critique <i>Albert O. Hirschman and Richard M. Bird</i>	6
The Project Cycle <i>Warren C. Baum</i>	13
Project Planning and Implementation in Developing Countries <i>Dennis A. Rondinelli</i>	21
Program Management and the Federal Evaluator <i>Pamela Horst, Joe N. Nay, John W. Scarlon, and Joseph S. Wholey</i>	28
Formulating the Question and Measuring the Answer <i>Carol H. Weiss</i>	36
What Can We Actually Get from Program Evaluation <i>Joseph S. Wholey</i>	55
Principles and Methods of Program Evaluation <i>Herbert D. Turner</i>	61
Appendix The Logical Framework	67

Contributors

- WARREN C. BAUM** Associate Director of Projects, the World Bank.
- RICHARD M. BIRD** Professor, Department of Policy Analysis, the University of Toronto.
- ALBERT O. HIRSCHMAN** Professor of Economics, Harvard University.
- PAMELA HORST** Senior Research Assistant, The Urban Institute.
- JOE N. NAY** Senior Research Associate, The Urban Institute.
- DENNIS RONDINELLI** Director of Graduate Planning Program and Professor at the Maxwell Graduate School of Citizenship and Public Affairs, Syracuse University.
- JOHN W. SCANLON** Senior Research Associate, The Urban Institute.
- RUSSELL STOUT, JR.** Associate Director of PASITAM and Lecturer, School for Public and Environmental Affairs, Indiana University.
- HERBERT D. TURNER** Chief of Program Evaluation, United States Agency for International Development.
- CAROL H. WEISS** Social Researcher, Bureau of Applied Sciences and Social Research, Columbia University.
- JOSEPH S. WHOLEY** Principal Research Associate, The Urban Institute.

Project Design and the Logical Framework

This is a module for the study of project design and of a project-level design tool, the *Logical Framework*, used by the US Agency for International Development (AID).

The term *project* needs some clarification. The word is applied to a variety of activities. In essence, a project is a specialized arrangement for addressing a particular problem or set of problems. A project typically has one or more explicit *goals*, operates within a *specified time-frame*, and includes a clear-cut *financial plan* and a *discrete organizational format*.

Projects may fit this generic definition in a number of ways. Some are simply arrangements of managerial convenience for structuring the work of organizations with conventional, continuing programs. Other projects package activities related to, but outside, the normal competence of ongoing organizations, e.g., the installation of some new technology which will then be operated by the receiving agency. In the field of development, the project device is widely used to undertake activities which lie outside the initial capabilities of ongoing organizations.

Projects are not the only types of development efforts, but the project approach is one basic way in which particular development problems are attacked. In fact, project design involves defining the problems to which projects are responses. Ultimately, "the problem" is that set of conditions which the project aims to change for the better.

The project approach has a number of appeals. It offers both focus and boundaries—focus on specified aims, inputs, organization, methods, and costs. Projects serve as "units" to be compared, analyzed, aggregated, controlled, correlated, and even matrixed. They can also be insulated to spread risk and to minimize consequences of failure. In a way, projects reflect an

organizational concern with "control," since they offer the promise of being controllable activities. In LDCs, they have the added appeal of insulation from existing organizations and their limitations.

For years, individual projects were the primary instruments of development. Results were mixed. In the sixties, some massive "projects" did little to eliminate poverty and deprivation in LDCs and, in some cases, actually worsened conditions. Over the years, development assistance and the character (or content) of project-level activity has changed. Projects are now viewed as building blocks of broader "programmatic" concerns in development. This complicates project design by reducing the self-containment of projects. The intended tradeoff is a greater ability to monitor and measure project impact upon broad programmatic goals.

Hirschman and Bird, in "Foreign Aid: A Critique," take a critical look at this shift from a "project" to "program" approach in development assistance. Their argument is a grand overview of the problems of aid administration. It examines the link between project purpose and overall programmatic goals, and leaves some things to be clarified.

For example, we speak of "project management" and "project purpose," but the general approach to development funding is no longer limited to specific projects as it was in the days of dam-building and massive resettlement schemes. Bilateral US development assistance is now programmatic in that projects are tied to a country program (the Development Assistance Program, or DAP) or a sector program such as Rural/Agricultural Development. Disaggregating a program into finite and measurable projects may help assure results. It also poses practical questions—and problems—about the relation of project purpose to

program goals. Project designers and managers must deal with these matters in the complex, interrelated process of project design, implementation, and evaluation.

The Hirschman-Bird article deserves careful reading, not so much for its critique of foreign aid *per se*, but for its penetrating analysis of the differences between project and program approaches and of their potential effect on donor and recipient countries. (The argument is particularly germane in view of the current emphasis on human rights and the “poorest of the poor.”) All parties to an AID agreement must be willing to provide formal measurable evidence of progress in realizing program goals. Though Hirschman and Bird emphasize the donor’s *economic* influence, their argument is also valid with regard to social/ environment policies. There is no need to take a stand for or against either approach to appreciate the differences and the potential effects of each.

The pursuit of contemporary development assistance goals requires clear definitions of project purpose, readily communicated project outputs, and a sharp look at required inputs. Unlike the old “social engineering” project approach, today’s program perspective calls for the critical assessment of assumptions about project effects and the environment. We no longer casually assume a project will have positive social/ environmental impact, nor can we presume that implementation is a linear process equivalent to dam construction. By widening and deepening the aims of development efforts, we enlarge the “unknowns” which must be addressed and increase the need for analysis and for genuine collaboration between donor and recipient.

KEY PROBLEMS OF PROJECT DESIGN

The programmatic approach requires development agencies to disaggregate goals into subsets, or purposes, and to design projects to meet those purposes. The program provides the overall framework in which a project will operate, but the project defines how and when particular aims will be realized. Ideally, the two are inextricably and explicitly linked. Together, program goals and project purposes set a basis for development efforts which promises the most effective and efficient use of scarce resources.

Projects provide both focus and boundaries. By limiting the scope of an activity and sharpening the focus upon means-ends relationships, projects aim to make development efforts more manageable than looser approaches. Even so, projects do not escape the challenge

of uncertainty. In fact, the problems of uncertainty are greater than in the days of technology-centered efforts and turnkey projects because project and program goals are more complex and elusive.

Project design is often viewed as a linear process with sequential steps of design, implementation, and evaluation. Reality is not often so neat. In some projects (such as construction) where technology is relatively determinate, implementation and evaluation closely follow the original design, because that design can be informed by reliable knowledge of what does and what doesn’t work. We know we cannot build dams out of sand or big buildings out of paper. Please note that this does not resolve the esthetic qualities of buildings, bridges, or dams, but only the technical questions of whether they will stand and serve intended uses.

When the limiting conditions of high certainty do not hold, rational project design must be incremental and iterative. Even goal elements are subject to change. This leads to a set of very important problems: how to achieve authoritative contingent design while at the same time not write *everything* off as contingent and therefore not subject to compliance requirements.

These problems include: how to deal with unanticipated consequences; how to keep project goals to a manageable, realistic number; how to deal with crucial qualitative factors, e.g., the quality of management and information; how to build in learning to allow for potential changes in ends as well as means, and to minimize sterile formalistic post hoc evaluation; and how to do these things without taking forever in the pre-implementation phase of design.

Uncertainty is the crucial problem of development project design. When we move beyond the old combination of narrow aims and technological means, we cannot predict the future with much assurance. At best, we make good guesses. How do we distinguish the good guesses from the bad guesses? How do we increase the *probability* of making more good guesses than bad? There is no definitive answer to these questions—those who claim to have it are charlatans. But there are ways and means, tools and techniques, that can improve our guesses. You will find some instructive observations in this module.

THE DIMENSIONS OF PROJECT DESIGN

In practice, design is always palpable and particular. It is time and site specific, involving real people, unspecified assumptions, intuitive and pragmatic knowledge (art), and hope. There are no universal tools to resolve strategic design problems.

Yet there are *general* approaches and some general principles, rules, and techniques. We have classification systems, if not typologies. For example, Warren C. Baum, in "The Project Cycle," identifies categories of concerns which must be addressed in a certain domain of project design. Baum's piece was written for, and is heavily oriented toward, the World Bank's programming process; but it contains observations useful to any organization engaged in the highly uncertain task of administering development assistance.

All project designers face similar problems. Baum classifies these as:

1. Technical
2. Economic
3. Commercial
4. Financial
5. Managerial
6. Organizational

Of course, the World Bank is emphatically interested in loan feasibility. Foreign policy considerations, which are extremely important in bilateral programs, are not addressed. (This is changing as the US Congress becomes more and more concerned with multilateral agencies to which the US is a major contributor.) As a bilateral agency responsible to Congress, USAID adds to Baum's classification concerns with social soundness, programmatic salience, and not least, political appropriateness.

Baum stresses the interdependence of the six dimensions of project design and the cyclic nature of the process of project identification, implementation, and evaluation. In this iterative learning process, design tools assume major importance. For example, AID's Logical Framework stirs the designer to spell out his or her key assumptions and to consciously examine the links between these assumptions and the hierarchic arrangement of inputs, outputs, purpose, and goal. The logframe does not make the future more certain; it does make the designer identify the key factors of concern to him or her in an uncertain world, and map out a strategy which takes them into account. Intelligently used, the logframe becomes one means of negotiation and supervision during the course of project implementation. As Baum notes, from World Bank experience, both of these activities—negotiation and supervision—are important features of project management.

The generic concern in project design is *causality*. But cause-effect relationships are slippery and not as apparent as we would like. How, then, can we address

the various concrete problems of causality in project design? Dennis A. Rondinelli, in "Project Planning and Implementation in Developing Countries: An Introduction," provides a cogent explanation which is particularly relevant to the "New Directions" in US foreign assistance.

In the past, concerns with causality were limited to input-output relations. The effects of those outputs were assumed to be desirable. Now, partly because of the lessons of experience, the scope of the causal concerns in project design has been expanded at the very time that reliance upon "public works" activity has dwindled. Expanding the dimensions of project design has complicated the process.

An earlier focus on absolute economic growth (usually defined as an increase in GNP), which often benefited the few at the expense of the many, has shifted to the quest for a more equitable distribution of development benefits. Serving this new aim requires new knowledge, including techniques for determining if target populations are, or will be, positively affected by interventions. Thus, there is a growing interest in "evaluation."

Rondinelli, though not dealing specifically with evaluation, assesses the "Problems of Development Management." He finds that the shift in international development policy over the past ten years has given rise to problems which had not previously existed.

On the one hand, there is increasing need for flexible, adaptive management. On the other, there are invariable tendencies to centralize decision making and to deal with the slipperiness of the real world by defining it away. This sometimes amounts to designing for a world that does not exist. The way to avoid over-centralization and excessive reporting is to sharpen skills and tools available to the project designers so that a better "fit" between project and environment can be reached.

This in turn requires that we learn from experience. Such learning can only be achieved through a systematic (but not necessarily formalistic) means of stating our notions of causality and testing them—however roughly—by experience. We therefore try to build in evaluation in project design. How is this to be done? What are some of the important elements to be considered when designing projects with evaluation in mind?

EVALUATION AND THE PROBLEM OF LEARNING FROM EXPERIENCE

Avoiding the repetition of past errors and learning to

detect new ones is a difficult but not necessarily impossible organizational task. It depends upon an effective flow of information to feed the discriminating growth of knowledge. Much of the crucial information never becomes incorporated into the “vicarious” or formal written-down system of the organization. It remains kinetic—built into the minds of individuals. The dynamic, changing character of information flows in an organization is frequently misunderstood. So are the ways in which organizations absorb and act on data. “Organization learning” is much more than a formal measuring and paper process. Yet this fact does not negate the need for seeking and getting objective evidence of utility and questing after explicit explanations.

But in seeking these explanations, the tendency is to stress post hoc evaluation. This can never completely—or adequately—describe and explain the dynamic processes of design and implementation. The three interrelated facets of projects—design, implementation, and evaluation—cannot be handled separately. If the organization is to truly *learn*, evaluation begins the moment a project is conceived and ends only when all technical and political considerations have been satisfied by the project design. Post hoc evaluation is useful; but treating evaluation essentially as a separate, post hoc formal activity is not only unwise, it is insufficient and it nurtures an unrealistic separation between project management and evaluation.

Discussing evaluation invariably raises the question of political “contamination,” i.e., political considerations that override rational evaluation methodologies and designs. Nowhere is this more evident than in bilateral development assistance. Seemingly “irrational” projects are often quite rational when viewed, not simply as formal attempts to correct an existing situation, but rather as efforts to win political friends or influence political enemies. This is the frequent dilemma of project designers and project managers who are faced with the question of evaluation. On the one hand, there are social scientific approaches which purport to measure impact and effect. On the other, there are the demands of superiors (and other wielders of power) to create and carry out projects on the basis of a particular political interest or urgency. There is no easy way out of this situation; but project managers do learn how to make plans and explanations which satisfy organizational imperatives, even if they do not meet the stringent guidelines of social scientific research.

Pamela Horst et al., writing about “Program Management and the Federal Evaluator,” examine some of

the specific concerns in assessing social interventions. They do not resolve the problem of reconciling political rationality with formal rationality, but they do offer some useful knowledge about the latter. Their interest is in a “preassessment” tool to improve not only evaluation, but program management. This, of course, is the primary utility of the logframe. Evaluation is only one element, albeit a key one, in project design.

The most difficult step in program/project evaluation is getting designers to clearly identify means and ends. It is not always possible to do so, but we can take steps to clarify our own sense of what it is going to cost. It is important to keep in mind that project inputs include not only hardware and materials, but also the services necessary to get the project moving and, eventually, to become self-sustaining. If development projects are to succeed, interventions must be designed with host country environmental and technical capacities in mind. One mark of well-designed projects is a built-in basis for evaluation provided by explicit objective statements about key project assumptions and expected end-of-project status (EOPS).

The myriad problems of project design are extensively examined in Carol Weiss’s “Formulating the Question and Measuring the Answer,” which is part of a larger work (*Evaluation Research: Methods of Assessing Program Effectiveness*) on evaluation as a social science research activity. Our concern, however, is with project design, not research. That does not mean there aren’t things to be learned from researchers. Though the language may be different and the focus shifted, the questions posed by Weiss are pertinent to project design. Note in particular Weiss’s figure 1. It expresses the same concern with causal linkages which is emphasized in this essay. It cannot be stressed too often how important it is to consider these linkages when designing interventions in an existing order.

Quite often, we hear good managers described as “better at guessing.” Better guessers are not accidents of nature. They are persons who (sometimes implicitly and intuitively) have learned what works and what doesn’t in particular contexts—persons who are able to learn from past errors in order to improve future actions. Tools such as the logframe facilitate the shaping and communication of such experience-based knowledge. The tool is not the teacher, but it is a means to learning based upon the premise that cause-effect relationships can be postulated and tested. As Weiss points out, whether this can be done depends on the way we formulate the question and measure the answer. She stresses the importance of data and gives us some useful

pointers on how and when data can be collected. Weiss explains why it is useful to design with evaluation in mind. Better yet, she also helps to indicate how to go about it.

Joseph Wholey takes a more critical look at evaluation in "What Can We Actually Get from Program Evaluation?" Wholey brings a unique perspective to the problem: he approaches it both as an analyst and as a public official. He uses domestic programs as examples, but the general question applies to development projects as well. One striking thing about social interventions is not their uniqueness, but the range of similar problems which recur wherever such projects are carried out. The old question of "Who is doing what, for whom, and why?" applies whether the intervention is designed for Detroit or Dacca. Cultural milieu will differ, but these basic questions remain the same. If we are to profit at all from lessons of the past, knowledge about the answers to those questions must be available and transmissible.

THE LOGFRAME: A PROJECT/LEVEL DESIGN TOOL

A project is a synthetic thing—an artificial instrument of some developmental intent. Projects—worthy projects—are synthetic in another sense as well. They result from the effective synthesis of three kinds of activity: design, implementation, and evaluation.

How this synthesis is achieved will depend upon the concrete realities—and often the intangible uncertainties—of particular conditions. There are no rules and no tools by which to reduce project design and project implementation to completely non-discretionary routines. But there are rules and tools to help make projects manageable and to increase the

chances of success. One such tool is the Logical Framework, or logframe. Properly understood, and astutely used, it can facilitate the effective synthesis of design, implementation, and evaluation.

Herbert Turner addresses this broad concern in "Principles and Methods of Program Evaluation." Since the logframe is an AID tool and Turner is a member of that agency's staff, his language closely parallels that of the videotape which is part of this module. But note the similarity between Turner's concern with learning the lessons within AID and Weiss's general reference to the importance of such learning. The common needs in project design are knowledge and information.

When we *know* what the effects of our efforts are going to be, design is a relatively straightforward task. It is in this sense that the word is used in architecture and engineering. In social interventions, where knowledge is not so certain, design becomes a skillful blending of the known and the assumptions we must make about the unknown. A well-designed project is one which *works*, and works to serve intent, given the goals of the organization and the purpose of the project.

A frequent failing in projects is to *assume* away the lack of knowledge which often attends such endeavors. Yet there is no greater ignorance than to ignore the reality of ignorance. It is possible—indeed it is necessary—to address the problem of ignorance, of uncertainty, whenever it applies—as is so frequently the case when the aim is development. This module explores basic dimensions of the broad subject, *projects*, to focus upon key considerations in their design and to present a tool which is being used to synthesize the design, implementation, and evaluation of international development projects.

—Russell Stout, Jr.

Foreign Aid—A Critique

Albert O. Hirschman and Richard M. Bird

Foreign aid is as Janus-faced an institution as can be found. In a world of sovereign nations, rich and poor, it is an instrument of national policy which can be used by the rich to acquire influence and to increase their power. At the same time, foreign aid redistributes income from the rich to the poor and can thus serve to speed the latter's development.

While foreign aid might never have come into this world without its appeal to both national and transnational interests, it has also suffered from the resulting ambiguity about its "real" function. Unlike such pure power instruments as national military establishments, on the one hand, and overt redistribution mechanisms like the progressive income tax, on the other, foreign aid has never been firmly institutionalized. It has led a precarious existence, bolstered from time to time by cold war conflicts and then flagging again as immediate dangers passed, or the lack of a "domestic constituency" in the aid giving countries made itself more strongly felt, or certain unpleasant effects of aid giving became apparent. Lately, signals of a new crisis in aid giving have multiplied in the United States, there is disaffection and disenchantment as well in Western Europe and perhaps in the Soviet Union, and foreign aid is none too popular even in the recipient countries.

Current practice in foreign aid dates from the new principles introduced by the Kennedy Administration in the early sixties. Essentially, this country's doctrine moved at that time to embrace what has since become known as the "program approach" to foreign aid.

FROM PROJECT TO PROGRAM AID

The "project approach" had predominated through the fifties. The World Bank had been enjoined by its very statutes to extend loans only on the basis of specific projects (in transportation, power, agriculture, and so forth). The first activity of the United States in the field of aid to underdeveloped countries was technical (Point Four) assistance, which had necessarily a project content and which evolved naturally into capital assistance with a similar content. Important departures from this practice occurred in countries on the periphery of the Soviet bloc. To a number of these countries the United States extended massive military as well as economic assistance, with the latter usually being justified in terms of short-term import or budgetary requirements.

By 1960, criticism of the project approach was widespread. It was easy to show how development depended not on a few specific projects, but on an adequate overall investment effort, with respect to both aggregate size and composition, and how ill-designed fiscal, monetary, and foreign exchange policies could undercut the positive contribution of any individual project to economic growth. Economists further pointed out that the donor country was not really financing the project for which it was ostensibly granting funds, but rather the "marginal" project which the aid recipient would have just given up had he not been handed the additional resources for a project which he probably would have undertaken in any event. For these reasons, so it was argued, a look at the total spending

pattern of the recipient country is essential if one wishes to have some assurance that the aid funds are put to productive use. Finally, it was pointed out that project aid necessarily implies a series of biases and perverse incentives: it encourages the aid recipient to prepare large capital projects, to exaggerate the foreign exchange portion of the total cost of these projects, and to favor public infrastructure projects, which are most easily financed through loans or grants extended from one government to another for project purposes.

While these criticisms of the project approach all contributed to a change in the climate of expert opinion, another important reason for going from project to program aid was the desire to increase the level of aid to some key countries and to provide a solid institutional basis for aid giving at this higher level. Program aid was conceived as aid given "in bulk" on the basis of a general understanding between donor and recipient about the latter's development program and principal economic policies. (Other terms frequently used in connection with program aid are, in ascending order of euphemization, "leverage," "incentive programming," "making sure of self-help.")

As a result of what was then thought to be the model case of India, the accent was at first primarily on achieving agreement on the recipient's development plan, its size, priorities, and the resulting "resources gap" to be filled by aid in its various forms. But in most developing countries, development plans are primarily statements of intention. Further, even in the rare country with a highly operational development plan, the fulfillment of the plan's objectives would depend crucially, among other things, on "appropriate" fiscal, monetary, and other economic policies. In Latin America, moreover, program aid under the Alliance for Progress was to be forthcoming not only in connection with a broad agreement on economic development objectives, but was to be premised also on advances in social development that depended on the enactment and implementation of reforms in land tenure, income taxation, educational opportunity, and the like.

THE TWO AID BARGAINS COMPARED

The general idea of moving from the project to the program approach consisted, therefore, in laying the groundwork for a substantial and steady flow of aid through a meeting of minds between donor and recipient on central economic programs and policies of the recipient country.

When the matter is put in this way, the formidable difficulties of the program approach begin to appear.

No doubt, by moving the discussion between donor and recipient from where to build what kind of power station to fiscal, monetary, or agrarian reform policies, one is turning from peripheral to central issues of the recipient's decisions. But is that a good thing? We shall now argue that this move raises at least as many problems as it solves.

To facilitate the discussion, it is useful to attempt at this point a conceptual distinction between "pure" project and "pure" program aid. In the real world this distinction will of course be blurred, as these two archetypes of aid hardly ever appear in their pure forms. Hence it should be understood that our subsequent discussion does not cover every conceivable case of project or program aid, but tries to catch the essential difference between two diverse forms of aid giving. Moreover, we do not aim at extolling project aid, with whose problems and drawbacks we are familiar, but rather at bringing out, with project aid as a backdrop, the heretofore largely neglected political implications and side effects of program aid.

As a starting point for the discussion, we may imagine that aid is given in the form of a check drawn by the donor to the order of the recipient, without conditions or strings of any kind. This unconditional aid can then turn into conditional aid along two principal routes.

First, the donor can insist that the money be spent for certain specific purposes; the result is pure project aid as here defined. Second, the donor may require that the recipient country change some of its ways and policies as a condition for receiving the funds; this is our definition of pure program aid.

From the point of view of the recipient, there is a fundamental difference between the two bargains which may conceivably accompany the transfer of aid funds. Pure project aid forces the recipient country to substitute to some extent the donor's investment preferences for its own insofar as the use of the aid funds is concerned. As a result, the recipient country lands in a situation it senses as inferior to the one in which the same amount of aid would be available unconditionally. Nevertheless, the aid permits the country to achieve a position in which it is unequivocally better off than without aid, in the sense that more funds are forthcoming for some purposes while, generally speaking, investments that the country would have made in the absence of aid will not be curtailed. Thus, the conditions attached to pure project aid are not likely to arouse strong hostility in the recipient country and do not require the policy makers to sacrifice any

important objective which they would have been able to pursue in the absence of aid.

The situation changes significantly in the case of the bargain characteristic of pure program aid. The commitment a country undertakes in connection with this sort of aid is typically of the following kinds: to increase investment and decrease consumption; to increase the share of the private sector and decrease that of the public sector; to devalue the currency and thereby alter *relative* price relationships within the country; to throttle inflation and therefore strike a blow at the particular interest group whose turn it is to benefit from the next inflationary appropriation, credit expansion, or rise in prices or wages. There are others. In all these instances, compliance with the conditions attending program aid makes one group within the recipient country worse and another better off than before. The bargain preceding the granting of program aid also implies that the aid receiving government will alter its previous policy mix in such a way as to sacrifice in some measure objective A (say, a larger public sector) to objective B (say, growth).

Economists who have discussed the concept of community welfare have long been divided into two groups: those who deny, and those who affirm, that meaningful statements can be made about increases or decreases in collective welfare when, as a result of economic change, one group gains at the expense of another. There is no need for us to enter into this discussion, except to note that its protracted and stubborn nature testifies to the fundamental difference between the two situations that we have just described. With pure project aid, the recipient government can achieve all of its pre-aid objectives (plus some additional aid financed ones) and no group in the country need be any worse off. With the type of conditional program aid discussed here, the objectives of public policies will be reshuffled and some domestic group is likely to be hurt. Even though the total resources available to the country are increased through the aid, the hurt group cannot be directly compensated, at least in the short run, for its loss, by the very terms of the aid agreement.

We should mention here one particularly important way in which project aid shades off in the real world into program aid. When the project donor spends its funds on, say, a certain kind of power station, it will often have views, and will attempt to have them prevail, on such matters as accounting practices, power rates, administrative autonomy, and perhaps even pub-

lic versus private ownership of the utility. Project aid may then also involve policy changes that would hurt some groups or individuals. Even in this case, however, an important difference between project and program aid remains. Program aid is usually given in connection with changes in *central* economic policies of the recipient, whereas the policy changes the donor is liable to insist on in connection with project aid are germane to the construction and operation of the project and are therefore likely to be concerned with matters that are at some remove from the central policy concerns around which the more important group conflicts rage.

THE PROGRAM AID BARGAIN FURTHER CONSIDERED

It will, of course, be argued that whatever sacrifice is entailed in the policy changes required by the program aid bargain is more than fully compensated by the other side, namely the aid package itself. The fact that aid is accepted on these terms could be considered as evidence that there is nothing to worry about. After all, the recipient government could have refused aid (as Burma did in general, and Brazil and Colombia at one time or another, in connection with assistance from the International Monetary Fund) if it felt that the conditions were too harsh. But this application of the notion of revealed preference misses several points. In the first place, we were intent on showing the difference between two forms of conditional aid giving and on pointing out that the cost of obtaining aid is of a different nature in the two cases. In the second, it is a gross oversimplification to treat a government entering the program type of bargain on foreign aid like a consumer buying himself a bag of apples. Since aid, in this case, has as its counterpart a shift in national objectives and in the short term fortunes of different social groups, the bargain will be considered a bad one by the circles that value highly the objective that has been sacrificed and by those groups whose interests have been hurt. Hence, the very bargain that gives rise to program aid can and will be attacked directly by these circles and groups as being damaging to the national interest as they define it. Pure project aid is ordinarily immune to this kind of destabilizing side effect. Precisely for that reason, those who attack it will often resort to alleging that it is *impure* and carries some unavowed and excessive cost in terms of general economic or political policy commitments. In other words, to be effective, an attack

on project aid will attempt to prove that it is *really* the program type of aid.

The difference between a country or a country's government adopting certain changes in its central economic policies as a *quid pro quo* for aid and a consumer disbursing cash for a pound of apples goes deeper still. The program aid bargain is effective only if the government is genuinely convinced of the positive value of the policies it has adopted in conjunction with the aid—if there has been, that is, a genuine meeting of minds between donor and recipient about the economic policy measures conducive to development. It is as if the consumer were not only made to hand over the cash, but were asked to positively enjoy this act instead of sensing it as a cost. Moreover, the commitment of the recipient government is ordinarily not just to a single policy action, but to a *policy* that requires implementation through a practically infinite *series* of actions. A more correct comparison of the program aid bargain would therefore be to the decision of a person who joins the monastic orders; he does not usually consider his vows of poverty and chastity as a payment for the promise of eternal afterlife, but as something to be valued and perhaps enjoyed directly and independently of that promise.

One matter is already becoming clear: for the commitments entered into in the course of program aid negotiations to be faithfully adhered to, the recipient government ought to be so convinced of the correctness of the policies to which it commits itself that it would have followed these policies even without aid. Paradoxically, therefore, program aid is fully effective only when it does not achieve anything—when, that is, no *quid pro quo* (in the sense of a policy that would not have been undertaken in the absence of aid) is exacted as the price of aid. (It is ironic that, at least when it is effective, program aid is vulnerable to the very charge that has long been leveled—wrongly, we think—against project aid: namely, that one can never be sure that the project thus financed would not have been undertaken even in the absence of aid.)

In these situations, the donor would set himself the task of *rewarding* virtue (or rather, what he considers as such) where virtue appears of its own accord.

This is indeed a modest and manageable task, but it is also one that does not usually satisfy the donors. Precisely because the institutional basis and public opinion support of aid are so precarious in the donor country, the proponents and dispensers of aid have quite naturally felt compelled to make extraordinary

claims for what aid can accomplish. The most persistent of these claims has been that aid acts as a "catalyst." This term is meant to convey that aid makes the difference between stagnation (or perhaps deterioration) and vigorous economic growth of the recipient country, or between the recipient being hostile and being friendly to the donor country. To these traditional and exaggerated claims for aid, a new variant has been added by the program approach: namely, that aid, properly conditioned, makes the difference between the recipient following the "wrong" and adopting the "right" economic policies.

In this fashion, then, aid is not seen in the role of rewarding virtue, but in the role, infinitely more difficult, of bringing virtue into the world. Now the fact that aid is known to be available *if* certain policies are followed will sometimes serve to strengthen a domestic group genuinely and independently convinced of the correctness of these policies and it is therefore not inconceivable that aid will on occasion help this group to come to power. This is the ideal case in which program aid acts first as a catalyst and then achieves so complete a meeting of minds and so full a sharing of values and objectives between donor and recipient that from then on they will march hand in hand toward a better future.

We have on purpose drawn a caricature, for it is our conviction that this picture of program aid as a catalyst for virtuous policies belongs to the realm of rhapsodic fantasy. At best, situations in which aid helps virtue to triumph in this fashion are the exception rather than the rule. The normal case is far more prosaic: the knowledge that aid is available if certain policies are adopted serves to make these policies more attractive and less costly than they would otherwise be. These policies will therefore often be adopted by aid-hungry governments in spite of continuing doubts of the policy makers themselves, resistance from some quarters within the government, onslaught against the "deal" from the opposition, and general distaste for the whole procedure.

Naturally, doubts and reservations are not voiced at the moment of the aid compact; hence the delusion on the part of the donor that there has been a full meeting of minds. But soon after virtue has been "bought" through aid under these conditions, the reservations and resistances will find some expression—for example, through half-hearted implementation or sabotage of the policies agreed on—and relations between donor and recipient will promptly deteriorate as a result.

PROBLEMS ENCOUNTERED IN BUYING VIRTUE THROUGH AID

It may be argued that once a government has unequivocally committed itself to certain acts as a condition of receiving aid, there is a good chance it will convince itself that these acts are truly in the national interest, even though previously it may not have thought so. Psychologists have developed the theory of "cognitive dissonance" to analyze individual behavior in similar situations. The theory teaches that if a person engages in "discrepant behavior"—in acts, that is, which cannot be reconciled to what he considers his beliefs and values—he will attempt to reduce the resulting dissonance by changing his values in such a way that harmony is restored.

However, the theory also stresses another point that is crucial here: if the discrepant behavior *is induced by either carrot or stick*, there will be far less consequential value change than if the discrepant behavior occurs in some accidental, absentminded, or experimental fashion. If the behavior is rewarded (as it is, in our case, by the granting of aid), dissonance hardly arises, because in accounting for his behavior to himself, the actor has a ready explanation and excuse for the fact that he did something contrary to his principles, opinions, or preferences. (For the same reason, declarations of support for a cause against which one has previously fought are unlikely to change a subject's prior beliefs when such declarations are exacted under torture.) Therefore, the very act of rewarding policy changes through aid undermines the determination with which these changes will be carried out and makes backsliding and sabotage more likely.

These considerations explain why certain types of policy commitments on the part of aid receiving countries are more workable—and therefore have turned out to be more popular with the donors than others. The more workable and more popular commitments are precisely those that are highly visible, verifiable, measurable and, at their best, irreversible. One thinks of a revision of the customs tariff, of the imposition of credit restrictions in order to curb inflation, or most typically perhaps, of a devaluation. In the latter case, there would seem to be little possibility of backsliding or of second thoughts. Yet, while devaluation cannot be retracted, its intended effects can usually be frustrated by subsequent monetary, fiscal, and wage-price policies. Hence, even in the case of devaluation, a government which harbors a feeling that it has been pushed into an unwise policy can often administer an

"I-told-you-so" lesson to the donor just by not carrying out certain complementary policies after the devaluation.

In the case of other economic or social policies that sometimes have stood in the center of aid negotiations, the continued psychological resistance of the aid recipients to such policies after a formal compact has been sealed can manifest itself more directly and easily. Whether the aid negotiations were concerned with enlarging the private sector of the economy or with establishing the basis for a land reform, the commitments a government had undertaken in these areas can be rendered inoperative through bureaucratic harassment or through lack of administrative energy, respectively. The old Spanish colonial adage *se acata pero no se cumple* (one obeys but one does not comply) will thus be widely practiced once again, and properly so. A country which permits its key economic policies to be determined by this type of international negotiation finds itself in fact in a semicolonial situation and is likely to adopt all the time-honored methods of stealthy and indirect resistance appropriate to that situation.

THE HIDDEN COSTS OF PROGRAM AID

The resistance of the recipient country to some of the policy commitments it has underwritten in the course of the aid negotiations is not the whole story. The general unhappiness about having had its arm twisted can find other outlets than backsliding on these same commitments.

In a simple model of international relations we may assume that, for the sake of independence, self-respect, and defense against accusations of being a satellite, the government of B, a poor country, is determined to maintain a certain *average distance* from country A, a great power and a potential donor. Country B measures this distance along two dimensions, the extent to which it adopts economic policies suggested by A and the extent to which it takes A's position in the leading issues of international politics. Under these conditions, success on the part of the great power in having B "do the right thing" in economic policy will result in a strong urge on the part of B to compensate for this move in the direction of A by a move in the opposite direction in international politics. Only in this fashion can the desired average distance be maintained. That this model of international behavior is not completely unrealistic, in spite of its simplicity, can be shown by recalling a few episodes of the recent past: the attempt

of the Quadros government in Brazil to move in the direction of a strongly neutralist posture in international relations after having adopted economic policies long advocated by the United States and the International Monetary Fund; to some extent, Pakistan's *rapprochement* with China; and, lately, a number of "surprising" foreign policy positions taken by the present Indian government just after it had finally been so "reasonable" in its decisions on domestic economic policy.

In this manner, a "successful" program aid negotiation in the course of which the recipient agrees to a variety of economic policies suggested by the donor may well have hidden, though considerable, costs: first, a direct cost to the donor in terms of the loss of certain diplomatic and foreign policy supports he thought doubly secure because of the aid extended; second, a serious loss of public support for the aid program in the donor country, as a result of what will be felt as ingratitude, hostility, and "irresponsible antics" on the part of the recipient. In this indirect fashion, the attempt at maximizing the productivity of aid by exercising "leverage" the risk of drying up the flow of aid at its very source.

OTHER FRICTIONS CREATED BY THE PROGRAM APPROACH

Our case can be further bolstered by important differences between project and program aid related to the diplomacy of the aid process. Consider first the donor's claim to have his advice taken seriously on the ground that he contributes substantial resources. This claim is strong in the case of projects, where the donor's contribution often amounts to one-half or more of the total cost of the project. It is much weaker in the case of program aid, for here the donor's contribution is measured against the recipient country's national product or, at best, its total investment or imports. In such comparisons, the aid effort is almost always likely to look disproportionately small in relation to the important changes in national economic policies that are being sought.

Next, we may examine the donor country's implied claim that its own judgment is superior to that of the recipient. In the case of projects financed by the donor, the justification and credibility of the claim is usually quite strong. The donor country is likely to know more about the construction of highways and power stations than the recipient, simply because it is economically more advanced and has specialized knowledge in the areas in which it stands ready to finance projects. When

it comes to appropriate economic policies to foster growth along with price stability and an acceptable distribution of income, the claim of the donor country to superiority is far more questionable. Frequently the donor country itself is far from having fully solved these very problems. Even if it has done better at them than the aid recipient, the applicability of its experience to the wholly different economic, social, historical, and political circumstances of another country must be much in doubt. The claim to superior knowledge is therefore fairly credible and innocuous in the case of project aid. It is not credible in the case of program aid—indeed, it is profoundly irritating.

The diplomacy of aid is even more directly involved in our final point. It is in the nature of the aid relationship that comparatively low level officials of the donor country are paired off in aid negotiations with high level officials of the recipient countries. This irksome difference in levels is far less pronounced in the case of project aid than program aid. In discussing the layout and specifications of a highway, an engineer of an aid mission or of the World Bank may perhaps exchange arguments at one point with the director of the highway agency of the aid receiving country. But the matters discussed in conjunction with program aid relate, as we have seen, to central economic policies and issues. Given the centralization of decision making and the thinness of the elite in the typical aid receiving country, these matters can ordinarily be decided only at the very top of the political structure, by the President and his Minister of Finance. And who are their counterparts around the negotiating table? At best, the director of the local aid mission and, usually, various mission staff members. In this way, program aid recreates a typical colonial situation in which the rulers of the recipient country have to deal as equals with, and often feel that they have to take orders from, persons who, within their own country, are miles away from the seat of power. There is no need to expand on the resentment created by this situation.

Since, in our opinion, the program approach overreaches itself when it attempts grandiosely to bring virtue into the world, the explicit or implicit conditioning of aid on changes in policies of the recipient countries should be avoided. This does not mean that the donor cannot make his opinions and preferences known; but it does imply that elaborate arrangements should be made to divorce the exchange of opinions about suitable economic policies from the actual aid giving process. The educational virtues of

12 / Hirschman-Bird

such discussions will be strengthened rather than weakened as a result. Finally, the donor should resist the temptation to measure "performance" of the recipient at frequent intervals by narrow quantitative

indicators, when by its very nature such performance can be assessed properly only over a relatively long period of time by a combination of quantitative information and qualitative judgment.

The Project Cycle

Warren C. Baum

If the question: "What does the World Bank do?" had to be answered in a few words, these words might be: "It lends for development projects." The financing of specific projects, carefully selected and prepared, thoroughly appraised, and closely supervised, is the Bank's central business. The Bank and, beginning in 1961 its soft-loan affiliate, the International Development Association (IDA) have since their inception made some 950 loans and credits* for a total amount of about US \$17 billion. Of these the overwhelming majority have been for specific projects such as schools, irrigation dams, power plants, and roads. In the fiscal year 1969, for example, 111 loans and credits (a record number) were signed, all but one of which were for individual projects. This emphasis on project lending derives from the Articles of Agreement of the Bank, which provide that loans made by the Bank shall, except in special circumstances, be for the purpose of specific projects of reconstruction or development.

Project lending is intended to ensure that Bank funds are invested in sound, productive projects with the purposes of contributing both to the borrowing country's capacity to repay and to the development of its economy. It is in the coincidence of these two purposes that the Bank's functions as an international financial institution merge with those that it has increasingly assumed as a development institution.

*Loans are made by the Bank at an annual interest rate which is currently 7 percent. IDA offers credits on concessionary terms of 50 years with a service charge of $\frac{3}{4}$ percent a year. The methods of project work are identical for the two institutions, and in this article all references to the Bank and to loans apply equally to IDA and to credits.

Even with the record amount of lending recently achieved, and the even larger amounts expected in the future, the Bank and IDA are and will be dealing with a relatively small number of project operations. These projects, characteristically, are large, usually costing in the millions of dollars. They are diverse, ranging from oil palm plantations in Papua/New Guinea to the construction of Japan's high-speed Tokaido railway line. No two projects are alike; each has its own particular history, and project lending has to be tailor-made to fit each one of them. Techniques and procedures of project lending are continually being re-examined and revised in a quest for excellence which is the challenge and reward of project work.

This work, as it is carried out in the Bank, is a continuous and self-sustaining cycle of activity, which runs through four principal stages. The first is identification of the project; the second its preparation, once identified, to the stage where it is ready for appraisal; the third is appraisal of the project and, assuming the appraisal is affirmative, the negotiation and signature of a loan for it; and the fourth is the supervision of the project in its construction and operating stages to make sure it achieves its objectives.

The reason that we consider this a cycle is that each stage not only grows out of the preceding ones, but leads into the subsequent ones, and we try to make it a self-renewing cycle so that new projects grow out of old ones in a continuous process.

IDENTIFICATION

The logical place to start is with identification. This is the process of identifying, in a preliminary way,

projects which are of high priority, which might be suitable for Bank financing, and which the borrower, the Bank, and the government are interested in considering. There are essentially three tests involved in the identification of a project. The first is whether the sector of the economy into which the project falls and the project itself are of high priority for development and are so recognized in the government's development plans. The second is whether, on prima facie grounds, the project seems to be feasible; that is, whether a technical solution to the problem to which the project is addressed can be found at a cost commensurate with the benefits to be expected. And the third test is whether the government is willing to support the project by financial and other means. This is because loans are made either to governments or government entities, or to private entities with the support and guarantee of the government.

How do we go about identifying projects? The answer is not a simple one. Just as there is a wide diversity of projects for which we lend, there is also a wide variety of ways in which projects come to the attention of the Bank, are identified, and are brought forward.

The most straightforward way of getting a new project is by a repeater loan to the same borrower, for subsequent stages of development of the same or a similar project. The Bank is now more than 20 years old and, as might be expected, we are repeating operations with a number of our borrowers. This is particularly so in the older fields of lending such as power and transportation. To give some outstanding examples, we have made ten loans and credits to the Indian Railways, to finance successive annual or biannual segments of their investment program. We have made seven loans to the Comisión Federal de Electricidad for various stages of expansion of the Mexican power system. So, a principal source of new projects is our continuing relations with borrowers, with new projects being identified and growing out of the supervision of existing projects.

I do not want to leave the impression that these repeater loans, because they are easy to identify, are necessarily easy to prepare and negotiate. Sometimes, this is so, and if the first project is going well, the second can be relatively straightforward. But sometimes the reverse is true. The first loan must be made somewhat on faith, with limited knowledge of the borrowing entity which may be a new institution established as part of the project, or with which we are just beginning to get acquainted. Various objectives are agreed upon and commitments entered into, but the project may not develop as hoped or expected. When

this happens it may take three, four, or even five years to make the second loan because the problems which arose under the first have to be straightened out and we must be reasonably satisfied with performance under it before we are prepared to make another one. So, while some of the repeater loans are easy, some are very difficult.

"Piggyback" Operations

There is a variant of the repeater loan, bearing the inelegant but descriptive title, which it acquired from the railway transportation field, of the "piggyback" operation. An important way of securing a self-sustaining project cycle is to include in a loan for a given project the funds for feasibility studies or for detailed engineering of subsequent projects. If, for instance, the Bank is financing the construction of a dam as the first stage of expansion of a power system, the loan may well include the studies necessary to identify and prepare the succeeding stages of the power program.

Missions

There are other ways in which projects are identified. The Bank sometimes sends a special mission to a country, to look into sectors in which we have not done business before. A team of specialists is sent on a preliminary reconnaissance of the sector to determine whether worthwhile projects can be identified. In agriculture and education we have formal relationships with other specialized agencies of the United Nations, which cover the identification and preparation of projects. In agriculture, the agreement is with the Food and Agriculture Organization (FAO), under which it staffs missions that perform identification work for the Bank and that also help to prepare those projects that have been identified. There is a large separate staff in the FAO working full time for the Bank on the identification and preparation of projects, with costs shared between the two institutions. The United Nations Educational, Scientific and Cultural Organization (UNESCO) performs similar services for the Bank in the field of education.

An indication of the importance that the Bank attaches to project identification may be seen in Africa, which is the newest area of the world in terms of Bank lending. Missions have been established in East and West Africa, with headquarters in Nairobi and Abidjan, respectively; between them, all of Africa south of the Sahara is covered. The terms of reference of these

missions are to help identify and prepare projects for Bank and IDA lending. They have succeeded in filling the pipelines with projects that are now reaching the lending stage in significant numbers.

Another important way of identifying projects is through the work of economic missions. The Bank periodically sends economic missions to its developing member countries for a variety of purposes, one of which is to review the major sectors of the economy in order to establish development priorities and to identify projects which might be suitable for lending.

Other Means

Finally, there are some projects which are identified without our help. These are usually brought forward by new borrowers interested in having Bank assistance, or by private business organizations. When, for example, an international steel or aluminum company is considering an investment to develop mineral resources in Africa, it may seek the Bank's involvement to obtain both additional capital and the security which the Bank's presence provides, and it will take the initiative in bringing such projects to our attention.

Usually, however, the work of the identification has to be done by the Bank itself. With the higher lending goals which Mr. McNamara has set forth, identification is becoming increasingly important as a means of filling the pipeline with projects for which loans may be made in succeeding years.

PREPARATION

Once a project has been identified, it enters what we call the project pipeline, and an extensive period of close collaboration between the Bank and the eventual borrower begins. This phase of the project cycle, because it depends so much on the nature of the project, the past experience of the borrower, the relationship of the Bank with the country and the borrower, and many other factors, is the most difficult to characterize. Again I should like to stress the diversity of projects and the fact that each has its own particular history.

Preparation covers all the steps necessary to bring a project to the point where its technical, economic, and financial feasibilities have been established and it is ready for appraisal. For a repeater loan where the first project has gone well, the time can be as brief as a few months. At the other extreme, where we have had great difficulties with the borrower in identifying and helping to prepare a satisfactory project, or where we cannot lend to the country for other reasons (for example, until

it has improved the management of its economy), preparation may take five years or more. Some projects fall out during the process, and a loan or credit is never made.

The time consumed in the preparation process is a common source of complaint by governments in the developing countries which have an understandable desire to short-circuit project preparation and move directly from the stage of identification to the making of a loan. But the projects characteristically involve long-lived investments, and the time spent in ensuring that correct technical solutions have been found, in setting up the proper organization, in anticipating and dealing in advance with marketing and other problems, usually pays for itself many times over.

Preparation of a project involves decisions, based on technical judgments, about the site and location of the facility, and more broadly, on the appropriate technical features of the project itself. This is the stage in which soil, hydrographic, and hydrological investigations have to be undertaken, and the suitability and adequacy of the natural and other resources required for the project determined. It is also the stage where alternatives are systematically explored. If, for example, a borrower approaches us with a proposal to build a large dam for hydroelectric power, one of the first questions to be asked is whether a thermal plant could provide equivalent power at a lower cost. Or the proposed project may be for the construction of a four-lane asphalt highway, while a two-lane gravel road, improved in stages with the subsequent build-up of traffic, might in fact provide much higher returns. A new irrigation project has to be considered against the alternatives of improving existing irrigation facilities or of developing agriculture through less capital-intensive measures in rain-fed areas of the country. If the school buildings in an education project appear excessively ambitious, less elaborate structures might save funds that could be used for investments elsewhere in the education system. At the preparatory stage it is therefore essential that all the alternative ways of accomplishing the objectives be considered. This is done typically through feasibility studies, which provide the preliminary design and engineering of alternative technical solutions and make a preliminary assessment of the economic benefits of each.

For a revenue-earning enterprise, the preparatory stage also includes a preliminary analysis to make sure that its financial position will be satisfactory. And for new enterprises or old ones whose past performance has been inadequate, project preparation may entail a great

deal of work with the prospective borrower to bring about improvements in organization and management.

The Bank's Role

The formal responsibility for preparing projects is that of the borrower and not that of the Bank. This formal position was rigidly adhered to in the past for good reasons, one of which is the potential conflict of interest between the roles of the Bank in preparing projects and in lending for them. It was felt that if we had prepared a project, we could not be objective later in the appraisal of our own work. While the position is logical, it has not been able to withstand the pressure of events. Experience has demonstrated that we do not get enough good projects to appraise unless we are involved intimately in their identification and preparation. The result is that, instead of having an invisible dividing line, with identification and preparation of projects on one side and appraisal and supervision on the other, there is a continuing cycle in which the Bank is closely engaged at all stages. One of the benefits of this change of attitude is that, through better preparation, fewer projects are rejected at the appraisal stage, although the final version of the project may be quite different from its original conception.

In practice, the Bank does not usually do the preparatory work itself because we do not have the staff for it; what we do now undertake is the responsibility of ensuring that projects are adequately prepared. This again entails a variety of means. Some of our larger or more sophisticated borrowers are quite capable of preparing their own projects, particularly after the first or second loans when they have become accustomed to the Bank's requirements, and they are encouraged to do so. If they do not have adequate staff for the purpose, they may have a regular relationship with a consulting firm which prepares projects for them. If such a relationship does not exist we may urge the entity to get consultants to help prepare the feasibility studies, and guide it in how to go about the selection. The feasibility studies may be financed in various ways: by the borrower itself; out of the proceeds of an existing loan or credit ('piggyback'); occasionally by a Bank technical assistance grant or credit; or more frequently by a grant from the United Nations Development Program, for which the Bank may act as Executing Agent. In the case of FAO and UNESCO, and of the Bank's African missions, the staff which identifies the project also helps to prepare it, unless specialized studies are required.

There is a common misconception that the amount of work it takes to prepare a project varies directly with its

size. Small projects tend to be located in small countries, perhaps newly independent, and the work involved in preparing these projects is entirely out of proportion to their size. So, if anything, there is an inverse relationship between the amount of money we lend for a project, and the effort expended in identifying and preparing it.

APPRAISAL

Let us move on in the project cycle to the next stage, that of appraisal. We will assume that the project has been identified, that it has been well prepared, the studies and reports are complete, and the preliminary indications are favorable—in short, that the project is ready to go forward. The next step is what we call appraisal. Appraisal is perhaps the best known, but not necessarily the best understood, part of project work. In the early days of the Bank the appraisal mission might be our first contact with the project and the borrower, but now it usually comes at the end of a period—perhaps as long as several years—of close association. Nevertheless, it is an important stage in the cycle because it is a comprehensive and systematic review of all aspects of the project.

While consultants or specialized institutions are relied on to help prepare projects, the appraisal work is conducted almost exclusively by Bank staff. The appraisals are always made in the field; desk studies alone are not enough. How difficult the appraisal will be depends very much on how the project has been prepared. If the preparation has been done well, the appraisal can be relatively straightforward; if not, what was thought to be an appraisal mission then becomes in fact a project preparation mission, and a subsequent mission or missions may be necessary to complete the job. The appraisal can cover up to six aspects of a project: technical, economic, commercial, financial, managerial, and organizational.

Technical

On the *technical* side, we have to be sure that the alternatives have been adequately considered and the correct technical solutions found. This means the right combination of seeds, pesticides, and fertilizer for a particular crop-growing project, or the correct system of drainage for an irrigation project; if it is a road project, that the width of the road, the shoulders, and the thickness of the pavement are appropriate to the traffic; if a railroad construction project, that the best alignment has been found; if a port project, that the

design of the berths and the depth of dredging are correct for the kinds of vessels serving the port; if an education project, that the number and layout of classrooms and laboratories are suitable for the proposed input of students and the curriculum envisaged. The list could be extended indefinitely. All features of the project design, the cost estimates, and the construction schedule are re-examined and confirmed or revised as necessary.

Economic

Ensuring that the right technical solutions have been found is a major focus of the work of the appraisal team. It ties in very closely with the *economic* dimension of the appraisal. While the Bank has financed a nuclear power plant, exotic transport vehicles such as hydrofoils, and the most modern container port facilities, we are not looking for the most advanced technological solution but for the one which best fits the circumstances of the developing country. This sometimes involves scaling down ambitious schemes that would be white elephants into more realistic and modest projects that would better serve the development of the country. Sometimes the reverse is true, and ways have to be found of enlarging the technical dimensions of a project, as, for example, by developing additional markets for its products, before the technical solution yields an adequate economic return. The cost-benefit analysis of different technical solutions to arrive at the one which gives the highest economic return is the responsibility of the economist on the appraisal mission. The analysis will often have been done in successive stages during project preparation, but the appraisal is the point where the final review and assessment are made.

While referring to the economic dimension, it might be worth taking a moment to describe how we view a project, since it helps to understand the role of the economist. The Bank takes a broad view of what a project is intended to be. We try to get the best relationship, at a particular moment of time, between the project, the sector of the economy into which it fits, and the development program of the country. First, we want to know that the sectors in which we are lending are of high priority to the economic development of the country; this is done primarily through the work of the economic missions which review the overall development strategy. Then we try to determine that, within each of the sectors of lending, the projects we are financing are of high priority; this determination is made by the sector studies or project identification missions or in the other ways that I have mentioned.

When it comes to the appraisal of the project, the appraisal report generally discusses the relationship between the project, the sector, and the economy. In transportation, for example, each appraisal report describes the transportation system as a whole and its importance to the country's economic development. If it is a highway project, the report will examine its relationship to the railway system, and vice versa. At the same time, we review transportation policies throughout the sector. If there appear to be public policies which are inimical to the growth of transportation, whether in highways or elsewhere, these will be explored at length in the appraisal and we will attempt in the negotiation of the loan to get improvements in these policies. In education, power, and telecommunication, the "project" as defined by the Bank may embrace the investment program of the whole sector, and in any case is closely related to the sector; in agriculture, which is more diversified and accounts for a much larger share of the country's economic activity, it is more difficult to fit individual projects into a strategy for the sector as a whole, but we are attempting to do so to an increasing extent.

Commercial

The third dimension of the appraisal is the *commercial*, which is of particular importance to revenue-earning enterprises. Commercial considerations cover all arrangements for buying and selling under the project. The Bank seeks to ensure that the procedures used to procure the goods and services required for the project will give the borrower the best value for its money; in general, this implies the use of international competitive bidding, which I shall refer to again later. The commercial appraisal also includes an evaluation of the market demand for the output of the project, the adequacy of marketing channels, and the supply of raw materials, labor, and other resources required for the project.

Financial

The *financial* dimension is, of course, closely related to the commercial. The review of a revenue-earning enterprise is very comprehensive, covering all the significant financial aspects, but for purposes of discussion these can be separated into two issues. One is our concern that there be sufficient funds for the construction of the project. It is not the practice of the Bank or IDA to lend all the money that is required for a project; we may lend as little as 10 percent or 20

18 / The Project Cycle

percent, or as much as 70 percent or 80 percent, but never 100 percent.

The borrower must therefore provide part of the funds. In large projects, the borrower may not put up all of the remainder itself, but may go to other sources of finance such as the national agencies of capital-exporting countries. An important aspect of the appraisal may be to ensure that there is a financing plan which will make sufficient funds available to construct the project on schedule. When the borrower is a government that is known to have difficulty in raising local revenues, special arrangements may be proposed, such as advance appropriations or the earmarking of tax proceeds, to secure the necessary funds on time.

The other financial issue is whether the enterprise will be able to meet all its financial obligations when it is in operation. Since we typically lend directly to the enterprise that is carrying out the project, we are naturally concerned in the first instance that it be able to service its debt to the Bank. But we are also concerned that it be able to meet all its financial obligations, that it have adequate working capital, and that it generate enough funds from its internal resources to finance a reasonable proportion of its future capital requirements. So a detailed review is made of the finances of the enterprise, with projections of the balance sheet, the income statement, and the cash flow. Where the existing financial accounts are inadequate, a new accounting system may have to be established as a condition of Bank lending.

The financial review often highlights the need to adjust the structure, and particularly the level, of prices charged by the enterprise. Whether or not they are publicly owned, the enterprises financed by the Bank generally provide basic services and come under close public scrutiny. Because it may wish to subsidize the cost of such services to the consuming public as a matter of policy, or at times perhaps simply as the line of least resistance, the government may be reluctant to approve the increases in prices necessary to make the enterprise financially viable. But financial viability is a *sine qua non* of Bank lending to revenue-earning enterprises, and the question of rate adjustments may prove to be critical to the appraisal.

Managerial

The fifth dimension of project appraisal is the *managerial* one. Here we are concerned both with the adequacy of the top management to direct the construction of the project and manage it thereafter, and also with the adequacy of the staffing at all levels of the

enterprise or organization. In some countries this becomes a major concern. At times the issue is one of redundant staff, as when a national railway is forced to cut back its services in the face of road competition, and has to work out a schedule for orderly reductions in staff. At times there is an opposite concern that the local staff is insufficient, either in number or in quality, to fill key positions in the hierarchy, so that external assistance has to be provided until local managers can be trained. The proper balance at any particular time between the use of consultants and foreign advisors on the one hand, and of local staff and management on the other, is a delicate one. Consultation between the borrower and the Bank on key managerial appointments is now a standard loan provision.

Organizational

The last of the six dimensions, the *organizational* one, is concerned with the administrative structure of the enterprise or agency carrying out the project. A revenue-earning enterprise that is publicly owned should, in our view, have an adequate degree of autonomy to administer its affairs. It should be free from political interference and from the rigidity that is sometimes inherent in agencies administered directly by the government. So while the degree of autonomy is a principal question examined by the staff member concerned with organization, our interest goes beyond this to make sure that the whole internal structure, the chain of command, the way in which departments are organized, the flow of decisions, the allocation of responsibilities within the organization are reasonably efficient.

This broader concern applies not only to revenue-earning enterprises, but to any organization to which we lend. Thus, a reorganization of the highway department has been part of a number of loans or credits for the construction of highways. In our agricultural lending, the reform of existing institutions, or the establishment of new ones, for example to carry out a livestock project, is increasingly the rule.

These are some of the principal types of issues that arise in the appraisal process. We are continually trying to improve our appraisal procedures and techniques. Systems analysis has been introduced into the choice of alternatives. Risk and probability analyses are increasingly employed in the cost-benefit work, to deal with the uncertainty attached to many of the project estimates. Shadow prices are being applied where true economic costs are not accurately reflected in market prices. We also try to keep abreast of new technology

and new techniques to see that they are brought into use as rapidly as circumstances warrant. The scope and content of the appraisal process are thus evolving over time, against the background of related developments in technology, economics, and finance.

We have described how projects are identified, prepared, and appraised in the field. The appraisal mission returns, it writes its report, and the report is carefully reviewed, and reviewed again within the Bank. When it has finally been approved negotiators are invited.

NEGOTIATIONS

Negotiations are the stage at which the Bank and the borrower endeavor to agree on the various measures that the appraisal report has shown to be necessary to ensure the success of the project; these agreements are then converted into legal obligations, set out in the loan documents. To illustrate: let us suppose that the revenues of public utility are insufficient because there has been uneconomic pricing of electricity, and we have agreed with the borrower that, in order that the utility may earn an adequate rate of return and to finance a reasonable proportion of its new investments, prices must be increased by, say, 20 percent immediately and 10 percent in two years' time. There will be a financial covenant, decided upon during negotiations, which defines the overall financial objectives of the utility and specifies the necessary rate of return and the timing of the rate increases. The managerial, organization, commercial aspects—in fact all of the issues that have been raised prior to and during the appraisal—are dealt with in the loan documents. For example, if it is concluded that a road should have only a gravel surface rather than asphalt paving, the project description will stipulate this. If consultants are required in connection with some part of the project, the borrower will agree to recruit and maintain consultants satisfactory to the Bank. Thus, the negotiations and the drafting of the legal documents are an essential part of the process of ensuring that the findings of the appraisal are translated into actions that are agreed with the borrower and will be implemented on an acceptable schedule.

After negotiations, the appraisal report is updated to reflect the agreements reached during negotiations, and the revised report, together with related documents, is given to the Bank's Executive Directors. If the Executive Directors approve the operation, the loan or credit is then signed in a simple ceremony which marks the end of one stage of the cycle and the beginning of another.

SUPERVISION

The final stage in the life cycle of an individual project is its supervision during the period of construction and subsequent operation. The purpose of such supervision is to ensure that the project is executed as planned, or modified in the light of changing circumstances, so that the development objectives are achieved.

Supervision is the least glamorous part of project work. Once the ceremony attendant upon the signing of a loan or credit has passed, the attention of planners and policy makers shifts to the new projects that are coming along; this attitude is understandable and it is reinforced by the fact that many months may elapse before the "old" project begins to yield tangible results. Nevertheless, supervision is an essential stage in the project process. It is obvious that however well a project has been identified, prepared, appraised, and negotiated, it is only when it is properly executed that the development benefits are realized. Moreover, development is by definition a dynamic process, and the circumstances under which the project is implemented may not coincide in all respects with those envisaged at the time of appraisal. An important part of the supervision process may therefore be to adapt the project to these changing circumstances, and in this sense "supervision" is a misnomer which does not do justice to the work involved. It is for these reasons that we have concluded that adequate supervision should have first priority in the assignment of project personnel.

As with all other stages, project supervision takes place in a variety of ways. During negotiations we will have agreed on a schedule of progress reports which the borrower is to prepare and which cover such questions as the physical execution of the project, its costs, and, if it is a revenue-earning enterprise, the financial status of the borrower. These reports are generally submitted at three-month to six-month intervals, and they may be prepared by the borrower or by its consultants. They are reviewed by the Bank staff in Washington, the same team which has been involved in all the earlier stages of the project cycle, and problems brought to light by the reports are handled in the first instance by correspondence.

A principal feature of project supervision covers the procurement of the goods and services that are financed under the loan or credit. One of the most important roles of the Bank is to see that this procurement is carried out in accordance with guidelines that have been estab-

lished for this purpose and that are published in a booklet which is widely circulated. These guidelines are designed to ensure that the works are contracted or the equipment purchased in the most economical manner. The principal device that has been developed over a considerable period of time to this end is international competitive bidding among qualified contractors or manufacturers. The guidelines give detailed ground rules about how this international competitive bidding is to take place. Seeing that these rules are observed in practice—and a single loan or credit may involve anywhere from one to several hundred individual contracts, all of which may have to be reviewed by our staff—becomes a time-consuming job and one that we take very seriously. Sometimes it is relatively straightforward and routine; on other occasions it becomes a major issue, as, for example, in the case of a telecommunications or power project where orders are being placed in amounts of \$5 million, \$10 million, or \$15 million; and it may come to a very close choice among several international suppliers as to which is the lowest evaluated bid. I should make it clear that it is the borrower, not the Bank, that is responsible for evaluating the tenders or bidding documents. The Bank's role is to review the borrower's work to make sure that it is done properly and that the rules have been observed. Our approval is required before the contract can be placed. And you can be sure that, if there is a controversy, it is called to the Bank's attention at an early stage.

The tasks that I have been describing so far are the ones that are done primarily in headquarters. We also send supervision missions to the field to see how the project is going forward and what difficulties are being encountered in its execution. The frequency of these visits varies with the nature of the project. On the average, we send a supervision mission every nine months to active projects. This is not as often as we would like, and we are taking measures to increase the frequency of supervision visits. Ideally, if we could get supervision missions into the field two or three times a year, we would be able to work more closely with our borrowers to deal with problems expeditiously and to see that the project is going forward quickly and achieving the best results. In the periodic reviews of our supervision work some projects (currently about 10 percent of the total) are classified in a special "problem" category, and these are watched with particular care and receive more frequent visits.

Supervision is a process of learning by experience. The lessons drawn are used not only to modify the

on-going projects as necessary, but also in the formulation of similar types of projects in the future. In the supervision process the results of the Bank projects are thus being continuously evaluated. This is not, in our view, a substitute for a full-scale review of selected projects, once they have been in operation for a sufficient number of years, to measure their actual results against the original expectations. The Bank has completed or has in process several such reviews, and more are contemplated on a regular schedule. We have also collaborated with various outside groups which have conducted independent studies of Bank projects.

The supervision mission has among its tasks the review of the preparatory studies of future projects included (in "piggyback" fashion) in the loan or credit. It also provides the continuing relation with our borrowers from which new projects are identified and brought forward. In this way, supervision leads to identification and preparation, and the project cycle renews itself.

THE CYCLE REVIEWED

Summed up, here are the main points that I have tried to make:

1. the relatively small, but growing number of operations in which the World Bank engages each year;
2. the wide variety of projects for which we make loans or credits;
3. the wide variety of paths that they follow in entering the pipeline and moving through it;
4. the broad scope of our interest from development program to sector to individual project;
5. the importance of project supervision;
6. and particularly the close association that has grown up between the Bank, the government, and the borrower at all stages of the project cycle.

We have found that to help prepare and execute projects that will make a useful contribution to development we must get into the project cycle at the earliest stage and remain with it continuously.

Despite the differences that inevitably arise on occasion when difficult issues must be decided, the relations that have developed between the Bank and most of our borrowers in this continuing process have become quite good. They have come to appreciate that the Bank has no ax to grind, that we are in business to assist their development by lending for well-conceived and executed projects, and that this is our primary, indeed our only, interest in project work.

Project Planning and Implementation in Developing Countries

An Introduction

Dennis A. Rondinelli

International development policy has changed drastically over the past decade, and the principal instrument of capital and technical aid—the development project—is emerging to play a new and more crucial role in planned economic growth. Projects have become primary means of transferring assistance and of implementing capital investment plans, and their management now plays a central role in development administration. The United Nations observes that in developing countries most “administrators are more directly concerned with program and project administration than with other, more generic aspects of public administration.”¹

Both governments and assistance agencies have come to realize that without successful project implementation, plans are no more than wishes and developing economies stagnate or regress. Yet, traditional approaches to national planning based on macroeconomic models and econometric analysis proved less than successful in generating growth and modernization and narrow approaches to project lending, focused primarily on low-risk, “bankable” investment in physical infrastructure and heavy industry, were no more effective in generating economic progress, ameliorating poverty or promoting social equity.

Even where growth occurred, the spread effects in the national economy have been minimal. The World Bank’s operations report on Colombia, for instance, a country in which it has had a long history of extensive economic investment, notes the uneven distribution of growth; projects increased opportunities for only a limited proportion of the population. Field evaluations conclude, in this case, that institutional and adminis-

trative weaknesses severely hampered the rate and distribution of development.² Unemployment, the International Labor Office reports, is now “chronic and intractable” in nearly every developing nation; neither national planning or project-by-project investment has come to grips with serious social problems. “Despite a decade of unprecedented increase in gross national product (GNP) of the developing countries,” Robert McNamara noted in his 1973 address to the World Bank Board of Governors, “the poorest segments of their population have received relatively little benefit. . . . Nearly 800 million individuals—40 percent out of a total of two billion—survive on incomes estimated (in US purchasing power) at 30 cents a day in conditions of malnutrition, illiteracy and squalor. They are suffering poverty in the absolute sense.”³

Thus international assistance agencies and planning ministries in less developed nations are exploring alternative agencies that emphasize a broader sectoral framework for investment that seeks to spread economic progress to the poorest segments of developing societies. Sectoral project management may become, over the next quarter century, the primary means of transforming development plans into instruments of action.⁴

Even a cursory examination of the past decade’s development plans reveals startling changes in rhetoric and strategy. A decade ago developing countries emphasized increased export production, expansion of heavy manufacturing, and creation of showpieces of modernization in major metropolitan areas. But the results have done little to relieve the burdens of populations living outside urban enclaves in developing

nations, to ease social tensions or to increase political stability in Asia, Latin America and Africa. After a decade of widespread discontent and sporadic revolution, even in developing nations that achieve the highest rates of economic growth, priorities have changed. Improvement in living conditions—in all its dimensions, and for a large majority of the population—now imbues the plans of nearly all developing countries. “No longer is maximum economic growth the singular apex of goals,” declares the Philippine national development plan. “Equally desired are maximum employment, promotion of social development and more equitable distribution of income and wealth.”⁵ Rural modernization is given priority equal to urban industrialization. Agricultural development and expansion of social services gain new priorities in long-range capital budgets. To achieve these goals, international assistance agencies and governments in developing nations have substantially shifted investment policies in two major directions.

SHIFT TO A SECTOR FRAMEWORK FOR PROJECT LENDING

International assistance agencies, dissatisfied with the progress of macro-economic planning and project-by-project investment strategies are making significant changes in their aid efforts. The new emphasis is on multi-purpose, integrated projects identified and selected through sector analysis. Sector analysis is becoming the primary link between national development planning and individual project identification and selection. In its broadest sense, sector analysis is the quantitative and qualitative assessment of problems and needs within a well-defined segment of the national economy. Although developing countries and assistance agencies use different definitions for sectors, three broad categories seem to have emerged:

1. *Directly productive elements of the national economy*—such as agriculture, industry, tourism, mining, and export trade.

2. *Physical infrastructure systems*—such as transportation, highways, and roads, electric power networks, irrigation, housing, water and sewerage, telecommunications, and public works; and,

3. *Social services programs*—such as population control, human resources development, education, health, employment expansion, and urban development.

Sector analysis attempts to define each segment of the economy concisely, identify its major components,

activities and subsectors, examine the gap between planned targets and actual achievements, define future problems and needs, and translate those needs into investment opportunities. It thus attempts to disaggregate broad national plans into specific, operational goals, providing a framework for project investment.

THE SHIFT TO GROWTH WITH EQUITY GOALS

A second and more pervasive change in international development policy is the shift from industrial and physical infrastructure investment to projects that promote basic social change for the poorest groups in emerging societies.

The World Bank, for instance, has expanded broadly the scope of its project lending, placing greater priority on agriculture, education, population planning, tourism, housing, urban water and sewerage facilities, and rural development. In population planning, the World Bank is beginning to fund “packages of projects”—those with a diversity of components such as small clinics, large hospitals, personnel training, equipment, and contraceptive delivery systems—as opposed to its traditional approach of limiting loans to construction of physical facilities. Projects may also combine components from different sectors: rural development, for example, may include not only road construction, utility, and agricultural elements, but also educational facilities, small health clinics, social services, credit, and marketing assistance.

Similarly, the US Agency for International Development (USAID) has sought, since 1973, to relate its aid more to expansion of specific sectors within less developed countries than to total national growth or to a disconnected series of individual projects. It now concentrates on a few major human problems in three key sectors—food and nutrition, population programming and health, and human resource development—and on selected development problems such as transportation, urban and regional development and science and technology, attempting to increase income redistribution and reduce unemployment.

The United Nations Development Program (UNDP) is also engaged in intensive analysis of social needs. The majority of UNDP technical assistance and preinvestment projects consists of multi-year ventures in such fields as resource exploration, education and training, agricultural and industrial development, public administration reform, and disease eradication.

And international agencies are pursuing new directions even with traditional infrastructure and industrial investment projects. "Projects in the form of independent installations are giving way to sets of interdependent capital using facilities," note Mason and Asher in their evaluation of World Bank policies. "The capital requirements of the entire power sector are considered rather than those of a single power plant. The evaluation of road networks or of complementary rail and road facilities is replacing that of isolated road or rail projects."⁶

CHANGING REQUIREMENTS FOR PROJECT MANAGEMENT

The emerging interest in the social impact of projects implies substantial changes in the process of investment selection and management. Identification, appraisal and execution will become immensely more complex. Appraisal and selection can no longer be based merely on financial returns; analysis must consider social, political, and institutional as well as economic change. The paucity of data for project appraisal in social sectors will make calculation of cost-benefit measures more difficult. Nonfinancial outputs will become increasingly important as the demand for social and behavioral as well as economic change receive higher priority in project evaluation. Activities and functions of project management will become more fragmented and difficult to control by the sponsoring organization. And as complexity, fragmentation, uncertainty and risk become greater, they will increase the interdependencies between project implementation units and their external environments.

To achieve the new development goals, international agencies, governments and private investors will have to overcome a myriad of management problems by creating new and innovative procedures for project coordination and execution. Innovation, however, must evolve from a clear understanding of current project management problems in less developed countries.

PROBLEMS OF DEVELOPMENT PROJECT MANAGEMENT

International lending agencies traditionally funded development activities on a project-by-project basis, primarily construction and "hardware" investments with high rates of return. Economy, efficiency and the debt capacity of the borrower were primary considerations in appraisal, which emphasized prudence to

ensure that proposals were "bankable." Little attempt was made to assess the long-range development impact of individual projects or to ensure that they were integrated into national and regional development plans. Responsibilities for planning and management were divided among external donors, suppliers of technical assistance, domestic lenders, indigenous national, provincial and local government agencies, consultants, and sometimes client groups. Project management activities—identification, preparation, appraisal, implementation and supervision—often were poorly coordinated. Cooperation was hindered by the lack of integrative mechanisms and by the scarcity of qualified technicians and administrators. Weaknesses of project management diminished the effects of investment on economic growth and the social impact on intended beneficiaries. Project investment in less developed nations, moreover, proved to be a long and complex process. As the volume of lending increased and the specialized organizational machinery within lending agencies grew, procedures became more complicated and the lag-time between project identification and completion expanded. Thus despite more than a quarter century of experience, international funding agencies and ministries in less developed countries still report serious defects in project implementation.⁷ Among them are:

1. *Problems of integrating project identification and selection with national government planning.* Few developing countries have successfully made national development plans operational using macro-economic models to coordinate project investments. A number of serious deficiencies arise in coordinating project identification and national planning, including the lack of strong political and administrative support for national plans among operating ministries responsible for project implementation; ineffective communication of plan goals to public and private organizations with investment resources; and the failure to specify projects required for plan implementation. National development plans are often stated in vague and amorphous language, lack capital budgets and sufficient information to operationalize macro-economic analysis techniques. The lack of adequately trained planners in central planning agencies and operating ministries compounds the problem.

2. *Ineffective procedures for defining and preparing projects.* In many developing nations, national planning agencies and operating ministries lack formal procedures for determining needs, identifying investment opportunities or preparing projects for domestic and

international funding. Among the most serious problems in this aspect of project management are the inadequate absorptive capacity of developing nations to finance, execute and operate many types of projects, inadequate assessment of overall needs or markets for goods and services, leading to overinvestment in some projects and underinvestment in others, and promotion of "pet projects" by individuals and groups within government agencies and international assistance organizations. Development patterns are often distorted through imposition of funding agency priorities on recipient governments. Long lag periods in the processing of project proposals by government agencies and funding organizations and the inadequate assessment of requests for continuation of existing projects, have also caused problems.

3. Deficiencies in project design and feasibility analysis. Developing nations lacking indigenous capability, rely heavily on funding institutions and expatriate consulting firms to design and analyze the feasibility of project proposals. Despite the increasing technical assistance provided to developing nations by aid agencies, serious deficiencies are still found in design and analysis.⁸ Among the problems are:

- a. Failure to relate projects to national development policies and priorities.
- b. Designs that are ineffective or inappropriate for local conditions.
- c. Insufficient analysis and comparison of alternative methods and technologies for attaining project objectives.
- d. Failure to clearly specify immediate and short-term project goals or to relate design to stated goals.
- e. Failure to integrate capital construction and physical infrastructure projects into larger systems or networks.
- f. Lack of contingency planning to meet emergencies or unanticipated delays.
- g. Failure to select adequate baseline data required for monitoring progress and performing post-evaluations.
- h. Failure to plan for policy and administrative changes necessary to implement projects, such as tax incentives, land reform, subsidies for private investment, and construction of related physical infrastructure.
- i. Lack of interaction among project designers and ultimate users, clients and beneficiaries.

4. Difficulties with appraisal and selection. Preliminary preparation and feasibility studies should provide

sufficient information to appraise and select the most desirable proposals. Although international agencies, especially the World Bank, have developed highly sophisticated techniques for project appraisal and selection, a number of problems have arisen in attempting to apply them. Many governments in developing nations complain of the difficulty of estimating true costs of capital, comparing alternative sets of projects, estimating returns on public investments using social costs, and accounting adequately for inflation, price increases and unanticipated expenses affecting overall cost. Governments are often unable to commit all potentially available investment resources to feasible projects, owing to ineffective budgeting and selection procedures. Projects are frequently selected on the basis of total financial resources available for investment rather than on potential development impact or administrative capacity to execute them. And where they are applied, financial and technical criteria are usually emphasized to the neglect of political and social criteria.

5. Problems of activation and start-up. Once a project is selected, the ability to implement operations plans is a crucial determinant of ultimate success or failure. But developing countries face particular difficulties in activating, organizing and implementing projects. Aid agencies delegate responsibility almost totally to loan recipients and although they provide some technical assistance and participate in monitoring and supervision, project implementation relies heavily on the administrative capabilities of government institutions. Among the most frequently encountered problems of start-up are:

- a. Delays in granting necessary approvals for project activation, and procedural and bureaucratic delays at the national government level.
- b. Corruption, inter-ministerial rivalries, conflicts among government agencies and private organizations, and lack of coordination in allocation of project resources.
- c. Underestimation of amortization obligation and resource demands of other on-going projects, leading to heavy additional unplanned borrowing.
- d. Shortages of adequately trained and competent project managers and technical personnel.
- e. Outdated accounting procedures and ineffective methods of budgeting.
- f. Assignment of too many projects to a limited number of government agencies resulting in over-extension of their organizational and financial resources.

g. Inadequate international agency field capability to provide required technical assistance for project activation and organization.

h. Selection of an inappropriate organization for project implementation.

6. *Ineffective project execution, operation and supervision.* Many developing nations simply lack the managerial skills and technical competence to execute and operate approved projects. Severe shortages of trained planners, administrators and technicians plague nearly every Third World country. As a result the execution, operation and supervision of development projects suffer from the following deficiencies:

a. Failure to redesign project upon discovery of unanticipated obstacles to implementation.

b. Failure to collect and process information required to indicate achievement of performance targets and to correct deviations from work schedules and specifications.

c. Difficulty in obtaining local resources and materials during construction, leading to delays and cost over-runs.

d. Insufficient capacity or incompetence of local contractors.

e. Excessive fragmentation of responsibility for implementing among government organizations and agencies.

f. Inadequate internal resource and work scheduling system and poor reporting and control procedures.

g. Inadequate supervision and monitoring by central government ministries, planning agencies and international funding organizations.

h. Inability to procure required resources, materials and supplies.

i. Failure to develop indigenous management skills by using projects as training operations.

j. Excessive use of expatriate managers in project implementation and failure to develop counterpart administrators.

k. High personnel turnover, poor personnel training, inadequate salary structures to attract and hold skilled indigenous and foreign managers.

7. *Deficiencies in the coordination of project activities.* Development projects are complex sets of activities that must be managed within an organizationally complex environment. Project management, even in the least developed societies, requires interaction with and coordination of a variety of local and international agencies, and public and private contractors and the success of a project hinges on the ability of the management to integrate their activities. Some of the most

critical problems arise from deficiencies in coordination, such as the failure of government agencies to provide supporting facilities and services required for project execution, the failure of one ministry to complete a project supplying inputs needed for a project being undertaken by another ministry, and political interference in the construction or operation of a project. Failure to plan for, anticipate and adjust to adverse political or social impacts of projects on local populations, and conflicts among government ministries over budgetary allocations required to complete projects, also inhibit efficient administration.

8. *Problems in evaluation, diffusion of project results and follow-up action.* Post-evaluation and creation of procedures for diffusing and adapting outputs have received relatively little attention. Only recently have lending institutions established formal project evaluation procedures. Planning for the diffusion of results and transfer of resources to operating agencies at completion is generally haphazard if performed at all. Follow-up action—both in the form of assistance to output users and in the identification of opportunities for further investment—is also generally overlooked. More specifically, problems include inadequate or inappropriate use of project outputs, limited demonstration and spread effects, and the failure to establish systems for adapting appropriate technologies used in project implementation to other development activities. A serious deficiency is the failure to train personnel for new projects or for transfer to operating units after the current project is completed. Projects, moreover, are rarely terminated expeditiously when evaluation indicates inability to complete them successfully. Failure to transfer outputs, excess resources and appropriate technologies to operating agencies diminishes the value of development projects. Excessive delays in submitting completion and evaluation reports in many developing nations makes post-evaluation and follow-up analysis difficult.

To catalog the problems of project management in developing nations is not to condemn their governments, international agencies, and private corporations but serves, rather, to underline the need for new and more effective management procedures if projects are to have greater impact on improving social and economic conditions in developing nations.

IMPROVING PROJECT MANAGEMENT CAPACITY

If these diverse and serious problems in project implementation are to be ameliorated, aid agencies, gov-

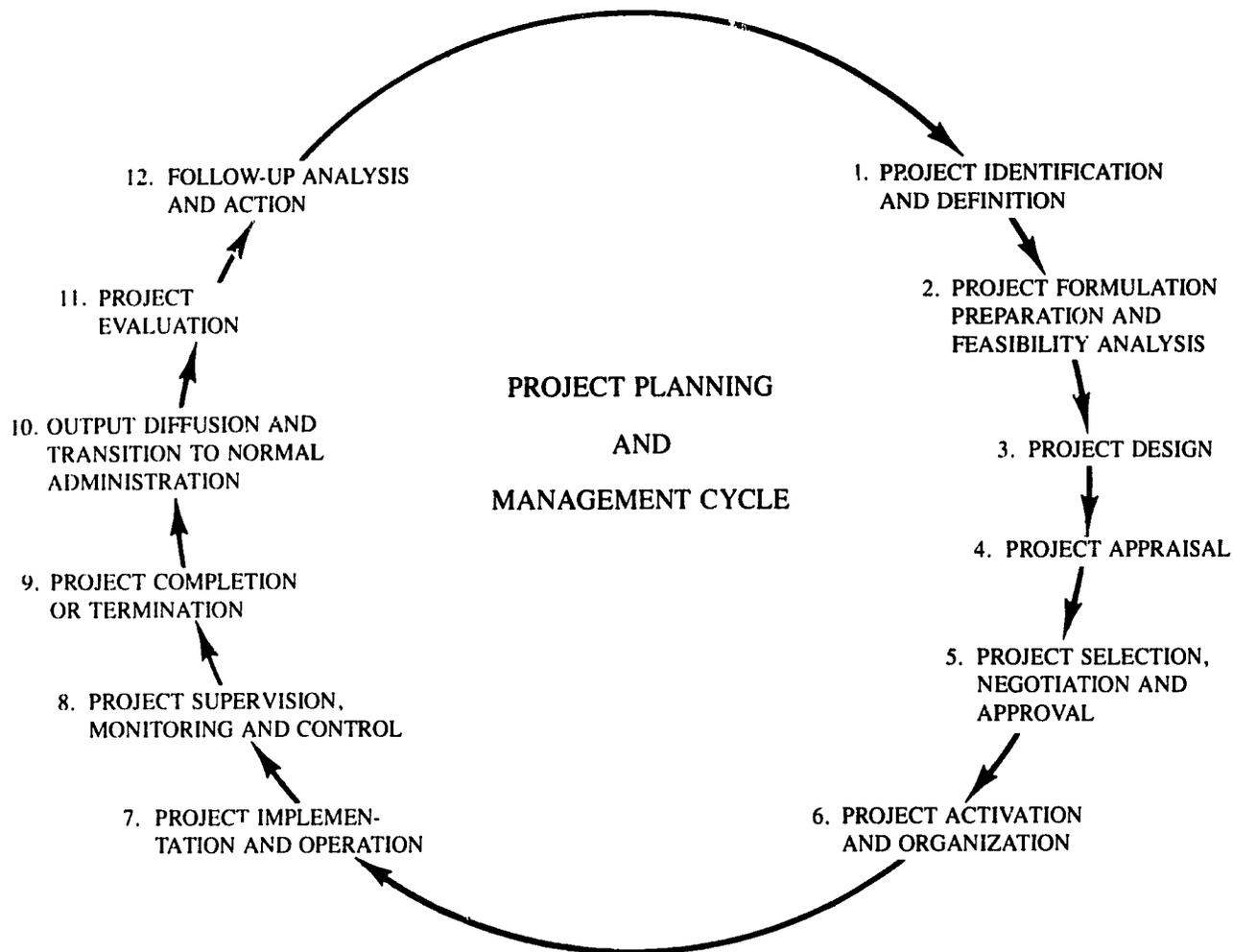


FIGURE 1

ernments of economically advanced nations, research and academic institutions and multinational corporations must join in with less developed nations to expand their administrative capacity to plan and implement social and economic investments. Expanding overall administrative capacity is a prerequisite to improving project implementation. "In one country after another," reports former World Bank official Albert Waterston, "it has been discovered that a major limitation in implementation projects and programs, and in operating them upon completion, is not financial resources, but administrative capacity."⁹ Expansion of administrative capacity requires training cadres of indigenous project administrators, experimenting with new organizational forms for project implementation, and formulating operational systems for project planning and execution.

TRAINING INDIGENOUS PROJECT MANAGERS

The need for highly skilled development administrators, with project management skills, is a recurring theme of international assistance agency evaluation reports. But traditional principles of public administration and management science are of little value in preparing administrators for the complex tasks of planning and executing integrated sectoral projects. Conventional public administration theories—advocating legalistic, centralized, regulatory procedures to ensure efficiency and economy in bureaucratic operations—are not adequate to deal with the complexities of social change in experimental development projects. Training that is available often takes a narrow focus, emphasizing project preparation and appraisal and directed primarily to high-level administrators or

staff analysts, mostly economists. Few training programs adequately expand the knowledge and skills of administrators in project organizing, resource mobilization, complex decision making, technical problem solving, the political aspects of coordination, and institution building, or in identification, acquisition, and utilization of non-economic resources.

EXPERIMENTING WITH NEW FORMS OF PROJECT ORGANIZATION

Expanding the number of indigenous project managers must be accompanied by experiments with new organizational forms for project implementation. "New managerial concepts and systems must be developed to organize and administer multi-purpose, multi-functional development programs which transcend departmental and ministerial jurisdictions," contends the United Nations Public Administration Division.¹⁰ At the national level, project selection and approval must be linked to planning, budgeting and allocation decisions. New organizational structures should all be tested to determine the most effective way of coordinating the diverse but related functions, within project implementation units.

FORMULATING AN OPERATIONAL FRAMEWORK FOR PROJECT PLANNING AND IMPLEMENTATION

Ameliorating the myriad of problems hindering project management will require a conceptual framework for testing new systems and procedures. Few government agencies in developing nations view a project as a related set of management activities; fewer plan and execute projects within an integrated management cycle. The most serious deficiencies in project management noted earlier arise from the failure to coordinate various stages of the project cycle. Thus governments seeking to improve their project management capabilities, must be prepared to perform at least twelve sets of administrative functions (see figure 1): identification and definition; formulation, preparation and feasibility analysis; design; appraisal; selection, negotiation and approval; activation and organization;

supervision, monitoring and control; completion or termination; output diffusion and transition to normal administration; evaluation; and follow-up analysis. These stages of the cycle are inextricably related. "It would be useful to visualize the phases in the life cycle of a project in terms of links in a chain," the United Nations advises, "and to think of project administration as an instrument for making these links of equal strength. Unless this is done the weakest link will affect the performance of the whole project."¹¹

Notes

1. See United Nations, *Administration of Development Programs and Projects: Some Major Issues* (New York: United Nations, 1971), quote from Preface.
2. International Bank for Reconstruction and Development, *Bank Operations in Colombia: An Evaluation* (Washington: World Bank, 1972). Unpublished report.
3. Robert S. McNamara, *Address to the Board of Governors* (Nairobi, Kenya: World Bank Group, 1973), p. 10.
4. Dennis A. Rondinelli, "Project Identification in Economic Development," *Journal of World Trade Law*, vol. 10, no. 3 (May-June 1976).
5. Republic of the Philippines, *Four Year Development Plan: 1974-1977* (Manila: National Economic and Development Authority, 1973), p. 307.
6. Edward S. Mason and Robert E. Asher, *The World Bank Since Bretton Woods* (Washington: The Brookings Institution, 1974), p. 249.
7. Dennis A. Rondinelli, "International Assistance Policy and Development Project Administration: The Impact of Imperious Rationality," *International Organization*, forthcoming; and Dennis A. Rondinelli, "Why Development Projects Fail: Problems of Project Management in Developing Countries," *Project Management Quarterly*, vol. 7, no. 1 (March 1976), pp. 10-15.
8. See Dennis A. Rondinelli, "International Requirements for Project Preparation: Aids or Obstacles to Development Planning?" *Journal of the American Institute of Planners*, vol. 42, no. 3 (July 1976).
9. Albert Waterston, *Development Planning: Lessons of Experience* (Baltimore: Johns Hopkins University Press, 1965), p. 249.
10. United Nations, Public Administration Division, *Public Administration in the Second United Nations Development Decade* (New York: United Nations, 1971), p. 5.
11. United Nations, Public Administration Division, "Some Aspects of Administration of Projects Within the Context of Development Planning," *Administrative Aspects of Planning* (New York: Economic Commission for Latin America, 1969), p. 402.

Program Management and the Federal Evaluator

Pamela Horst, Joe N. Nay, John W. Scanlon, and Joseph S. Wholey

In 1969, the Urban Institute completed an extensive study of federal evaluation and concluded that, "The most impressive finding about the evaluation of social programs in the federal government is that substantial work in this field is almost non-existent."¹ A limited resurvey of the field in 1972 revealed a quite different picture: funds committed to evaluation had mushroomed, many studies had been completed, and the use of large-scale social experimentation was increasing.²

This growth in evaluation has contributed information—often imperfect, sometimes incorrect—to today's arguments about the direction, method, and purpose of social programs. Without evaluation, many arguments would have remained at the level of polemic. There is no question that the presence of program evaluation has heightened the consciousness of federal program managers and policy makers to the fact that they may, from time to time, have to respond to queries about the effectiveness of their programs.

While evaluation has firmly established itself since 1969 in both the budget and the administrative rhetoric of the federal government, there is little evidence to show that evaluation generally leads to more effective social policies or programs. On the contrary, the experience to date strongly suggests that social programs have not been as effective as expected and have not improved in performance following evaluation. This situation can be phrased as a critical management problem which we see confronting government agencies:

Why have those in charge of programs and those who evaluate them not been able to join their efforts in a way that leads more frequently to significant improvements in program performance?

Having been able both to observe and to participate in the development of federal program evaluation, we have chosen here to raise three propositions about the root causes of the above problem. If these causes are the crucial ones and if we can come to understand their true impact, federal program management and evaluation stand on the edge of a period of increasing success. If not, and the causes of these weaknesses continue to be ignored, then evaluation, program management, the programs themselves, and those the programs are intended to serve will all continue to suffer.

This paper elaborates on why the three root causes, when they exist, block further improvement of many programs. The idea of a "preassessment" of program evaluability is introduced as one tool for improving both program management and program evaluation. We begin with a discussion of the conventional treatment of evaluation problems and then present an alternative diagnosis and prescription. Although much of the material presented is addressed to federal managers and federal evaluators, we believe the problems and solutions discussed also hold for state and local government.

APPARENT CAUSES OF EVALUATION PROBLEMS— AND AN ALTERNATIVE STATEMENT

Most reviews made to determine what causes programs and their evaluations to be ineffective include one or more of the following conclusions:³

1. Evaluations are not planned to support decision making.

2. The timing, format, and precision of evaluation studies are not geared to user needs.

3. Evaluation findings are not adequately communicated to decision makers.

4. Different evaluations of the same program are not comparable.

5. Evaluation fails to provide an accumulating, increasingly accurate body of evidence.

6. Evaluation studies often address unanswerable questions and produce inconclusive results.

The first three apparent causes deal with aspects of evaluation use. They occur at the interface between the producers of evaluations and the prospective users of evaluation.

The second three apparent causes deal with the methods used by the evaluators in assessing the interventions of the programs in society. They occur at the interface between the producers of evaluation and the program as it exists. They concern flaws in making measurements and comparisons and in drawing conclusions.

Our experience to date in studying the management problem—namely, the lack of significant improvement in program performance—and in watching various agencies attack the apparent causes of the problem has led us to conclude that these six statements largely refer to symptoms, rather than causes. We believe that the causes of the problem may more properly be described by one or more of the following three propositions concerning the program itself:

1. *Lack of Definition*: the problem addressed, the program intervention being made, the expected direct outcome of that intervention, or the expected impact on the overall society or on the problem addressed are not sufficiently well defined to be measurable.

2. *Lack of Clear Logic*: the logic of assumptions linking expenditure of resources, the implementation of a program intervention, the immediate outcome to be caused by that intervention, and the resulting impact are not specified or understood clearly enough to permit testing them.

3. *Lack of Management*: those in charge of the program lack the motivation, understanding, ability, or authority to act on evaluation measurements and comparisons of *actual* intervention activity, *actual* outcomes, and *actual* impact.

When one or more of these three propositions is true, both the problem (lack of significant improvement in program performance) and the six apparent causes listed earlier can easily occur. In cases where the first two propositions hold, an enormous range of

possibilities will present themselves as to which measurements and comparisons to make—with no criteria for making sound choices. In cases where the last proposition holds, even exceptionally high quality evaluation is not likely to be used well, if used at all. If a program suffers from one or more of these three flaws, there is a very low probability that evaluation information useful to program improvement can be produced. Thus the program may be “unevaluable” until the flaws are corrected.

A statement that the quality and value of evaluation are strongly affected by the degree to which these three conditions exist is not a startling finding. What has not been realized or acknowledged in the past, however, is that these three factors are not the responsibility of the evaluator. While the conventional apparent causes relate to how the evaluator does his job, these latter three propositions describe an organizational environment over which the evaluator typically has little control. Evaluators, more than any other group in an agency, will appear unable to complete their work successfully when these conditions exist, regardless of how they deal with the apparent causes.

WHY THE APPARENT CAUSES ARE SUSPECT

In the past few years, we have conducted a number of federal program evaluations and helped to develop evaluation planning systems for several federal agencies. In the course of our work, we have observed many attempts to treat the six apparent causes directly by improving the *use* and methodology of evaluation. These attempts include policy review and dissemination panels, letting contracts for methodology development, high level reviews of evaluation plans, task forces to select better questions for evaluation, better systems for collecting data, requiring program offices to submit advance descriptions of how evaluation findings will be used, and the tightening of contract selection and monitoring procedures to increase contractor responsiveness to agency needs. In some cases, the “solution” was reorganization: centralizing previously decentralized evaluation units. Since revenue sharing, talk of decentralizing a previously centralized evaluation office has gained popularity. In this case, the headquarters office would no longer be responsible for conducting national program evaluation, but instead would go into the business of building local evaluation capability.

As these proposed solutions were implemented, however, we have continued to talk with and work with

participants in the process from the assistant secretary level, through the program level, and on down to the recipients of services. We find that the management problem—that is, the lack of significant improvement of program performance—continues to exist and the same apparent causes continue to be cited, whether there are high or low quality evaluation efforts. Improvements in programs and in delivery of effective services remain far below the levels desired or expected. If the root causes of the problem lay within the evaluation process, we believe that these correctives would be showing some degree of success. Consequently this experience led us to search for alternative explanations and, finally, to the consideration of the three conditions stated above as root causes of the problem.

THE SOURCE OF THE PROBLEM

The significance of the proposed causes can best be understood by contrasting the nature of the intervention that the social programs of today attempt to make in the society at large with the principle types of program interventions attempted in the past. Many older, classical government activities involved program interventions whose nature was clearly defined and agreed upon and which were described in detail in a body of law or regulation (e.g., Social Security). The implementation of such activities was largely an act of administration of the laws and regulations. Evaluation of success or failure of the act of implementation was primarily a matter of assessing compliance with the guiding laws and regulations. Discretion was at a minimum (at least over the short term). Arguments might take place about whether *goals* were adequate, but the *details of the program intervention* were determined in advance.

In contrast, many new missions that the federal government has been called upon to undertake (e.g., lowering hard core unemployment) involve problems in which the proper program intervention mechanism is not well understood, or defined, or in some cases even known. Since in these cases no one knows exactly what detailed program intervention will be of value, greater management discretion is allowed and exercised. While in some cases research may be undertaken or experiments may be made to increase understanding, more typically a purportedly successful type of program intervention is simply put into place and an agency or bureau is charged with making it into a successful operation. In this case, evaluation is expected to report to those in charge of a program on

whether the use of discretion in choosing specific program intervention techniques was successful and perhaps to suggest modifications or alternatives.

The newer program areas are characterized by uncertainty and discretion: uncertainty as to the nature of the problem and what constitutes effective strategies of intervention, and discretion in how the problem and the intervention are defined and how the intervention is implemented. These conditions make sound and rapid evaluation all the more important to effective management. Consider, however, how today's program environments can disable evaluation through three factors: lack of definition, lack of a clear logic, and lack of management.

LACK OF DEFINITION

Examination of program legislation, regulations, policy manuals, plans, and budget to determine what a program intervention is can be very deceptive. What at first seems clear often evaporates when the test of measurability is applied. The language used turns out to be ambiguous precisely where it would have to be specific in order for evaluation to be useful. Three common forms of inadequate language are: the vaporous wish, local project packaging, and how-to-do-it rule making.

The vaporous wish is the eloquent but elusive language of goals put forward for most federal programs. Exactly what are the "unemployability," "alienation," "dependency," and "community tensions" some programs desire to reduce? How would one know when a program crossed the line, successfully converting "poor quality of life" into "adequate quality of life"? Would anyone recognize "improved mental health," "improved local capability," or "revitalized institutions"? The problems addressed by social programs are almost never stated so that institutions, people, or the relevant socioeconomic conditions could be classified according to the degree to which they are afflicted with a problem. It is very hard to propose a solution to a problem that is ill-defined or undefined. How much harder it is to evaluate the success of that proposed solution.

Next, there is the project packaging language which purports to describe the intervention activity to be planted in the field and the expected outcome for those directly served by that activity. As any experienced site visitor will attest, this language is often so annoyingly imprecise that it is difficult to tell what parts of a local

operation are under discussion and even harder to distinguish compliance or assess performance. For example, project characteristics prescribed in various program guidelines include: "coordinative mechanism," "integrated services," a "range of modalities," "extended career ladders," "accessibility of services," "continuity of care," "multi-disciplinary teams," "outreach capability," etc. Projects should produce "upgraded job skills," "increased cultural enrichment," "increased personal autonomy," "improved family cohesion," etc. Rarely are useful measures or norms for these activities and outcomes provided.

How-to-do-it rule making is the third kind of language that is commonly found. Here the terms are very concrete and specific. We find guidance on factors like the qualifications of project directors, the contents of affiliation agreements with other local agencies, reporting relationships, the use of consultants, and accounting practices. This guidance appears to be definite and all inclusive. Closer examination shows that it usually tells how to run the part of a project which does not deal directly with the intervention into society. Guidance for the part of the project which actually produces effects in society is not provided.

When these three forms of language predominate, the intervention activities in the field may be diverse indeed. Our experiences examining field operations indicate that program packaging is generally skin deep and that very different project activities and definitions of outcome often parade under the same assumed program names. An examination of 20 projects in the same program will often reveal 20 very different program intervention designs, different in activity and purpose. This means that the program activity and objective, as implemented in the field, cannot be defined on a common base of measurable terms. It is often difficult to find any consensus among federal level policy makers as to what the definitional base should be. This lack of a common framework can disable management and evaluation efforts alike.

It is becoming clearer that many federal social programs are simply envelopes for a large federal investment in a problem area. A program may be deceptive in the sense that it has enough content to allow it to be described in the media, lobbied into existence, and established as a federal effort—and yet the program interventions are not spelled out in any detail. Many program administrators over the last decade have essentially received a program envelope with only vaporous wishes and money inside. Although more detailed definition may not have been necessary in order to

spend the money, much more detailed definition is needed to evaluate the process and outcome.

If it is decided that certain programs should be further defined, who in any agency should be responsible for the tasks? It should not be left to the auditors or to the evaluators or to the information system people, because the choice of specific measurable definitions is not merely a technical task. The definition of what is to be measured in a program is central to policy making and program management. If there are many different ways to measure the problem a social program purports to influence, this often means that there are many different problems. For many programs, no one has yet exercised the prerogative of selecting which specific set of social ills the program is trying to cure or the methods of cure. Legislation or regulations rarely make this choice, and the choice has policy implications since it further specifies program intent and intervention. One of the major factors in shaping and directing a program is carefully selecting what the program is going to do. The failure to define measurable interventions, outcomes, and impact for a program is a major policy making defect. Those in charge of the agency and the program, rather than the evaluators, should have primary responsibility for program definitions.

LACK OF A CLEAR LOGIC OF TESTABLE ASSUMPTIONS

Even if the policy makers or program managers have provided measurable definitions, there still may not be unanimity within a federal social program about design or logic. As a result, different evaluation efforts are often based on *different* assumptions linking program intervention with immediate outcome and ultimate program impact. The measures and data collection instruments used are those that seem most reasonable to the evaluator. In this context it is easy to understand why evaluation findings are often noncomparable. When there is no carefully determined framework to guide the program, there is, of course, no such framework for evaluation studies. Nor is there a framework for systematically accumulating knowledge of program performance. In fact, it becomes unclear what program performance means.

Program assumptions might be as simple as that "the transfer of money to school districts will raise the reading level of disadvantaged students" or that "the training of the unemployed will lower unemployment." Often the board program charters from the Congress referred to earlier have caused clusters of competing

assumptions to grow up in many social programs. One set of assumptions may be used for arguments with friends, for instance, and another for arguments with enemies. This may be good politics, but it makes for difficult evaluation design, since evaluation design should relate to the information needed to validate, refute, or modify a set of operating assumptions.

Without an adequate description of the assumptions governing the intervention of a program into society, it is more likely that evaluators will be asked to address unanswerable questions far removed from the actual activities taking place. To take a quite reasonable example, a program office might insist on funding an evaluation to assess the relative effectiveness of different drug treatment modalities. The evaluator may then find that these modalities do not represent pure, mutually exclusive approaches which are replicated in multiple local settings. He is likely to find, on the contrary, that a "halfway house" or a "therapeutic community" in one locale bears no resemblance in operating assumptions to other: which go by the same name. After spending a lot of money, time, and effort, the evaluator will be forced to tell the agency what types of programs are really out there, rather than how successful they are, and also that the only way to test the effectiveness of alternative assumptions of treatment is to implement a program-level experiment, or introduce planned and enforced variations into the program design. Those in charge of the program may feel that the evaluator has once again failed to answer their questions. There are many examples of evaluations being mounted to answer questions which bear no relationship to the program activity actually taking place in the field. This counter-productive practice results from the failure of the agency to describe carefully the program assumptions so that they can be implemented and tested.

Summing up, even when the intervention, expected outcome, and impact are defined in measurable terms, the more subtle questions of the logic linking (a) program expenditures to production of the intervention, (b) intervention to outcome, and (c) outcome to impact on the problem must still be considered. The use of the word "logic" here is not meant to imply that the linking assumptions are loose or tight, valid or invalid, defensible or stupid. All that is implied is that a program in reality is based on an interrelated set of assumptions about what is believed to happen (and sometimes why) when money is spent and the intervention made. The absence of statements of these assumptions might be expected to cause a problem for both program managers and evaluators. The evaluators often notice the absence

first, however, because they must design tests of these assumptions. Tests cannot be designed for people who are unable to, or refuse to, state their assumptions.

Once again (as with measurable definitions) the statement of the logic of testable assumptions is a policy question, not one that should be decided by the evaluators. Evaluators should test the assumptions about what works. Those in charge should make the initial assumptions underlying the funding and operation of the program.

LACK OF MANAGEMENT

To get at the significance of lack of management, it is important to realize that evaluation is useful only if it is, in fact, a tool of management. A manager has a variety of tools to employ that include direction of his line management, planning, budgeting, audit and financial control, administration (for that part of his activity that can be clearly defined and where a law or set of rules is used to guide program implementation), policy analysis, and evaluation. Evaluation is needed principally in support of policy analysis and management discretion. Evaluation performs the same function for management that audit and control do for budgeting and that compliance checks do for administration.

One way of understanding the role of evaluation as a management tool is to explore how a "textbook manager" might use evaluation in attempts to improve program performance and then to contrast that with the way evaluation more frequently is used.

Evaluators and the "textbook manager" cooperate very well. When the policy decisions about program design are to be made, the evaluator asks the manager to specify the measurable definitions, the assumptions of the program linking these definitions, what kind of performance data would cause the manager to act, and the kinds of action the manager has the authority and willingness to implement. Armed with this guidance, the evaluator estimates the level of error associated with collecting the evidence, estimates the ranges of possible findings, and bounds the cost of the proposed evaluation. The evaluator is then equipped to provide a service not commonly rendered at present. He can advise management on the cost and feasibility of procuring evaluation evidence, and the manager can weigh these factors against the potential value of evidence for improving program performance. When the evaluation is finally commissioned, the evaluator has a clear basis for judging the best level of aggregation, precision, and delivery schedule because

he has a user for the proposed evaluation. Many market surveys and internal evaluations are conducted this way in industry. When this kind of rational planning occurs, one does not generate evaluation studies in search of users and uses.

The utility of social program evaluation depends at least in part upon defining the decision context as well as the program design. The "textbook manager" has already defined his program in measurable terms and has indicated what it purports to accomplish. If evaluation is to contribute to program improvement, there must be at least a few decision areas where the manager will rely on program performance feedback (measures of impact, outcome, intervention activities), as well as on political pressures, popular approaches, or his own hunches and beliefs. Else why buy evaluation at all? The "textbook manager" knows in advance and can specify what level of evidence will prompt him to act at all, or cause him to select among alternative actions. Further, he has the authority to act.

Return with us now to reality, where the typical government administrators live. These administrators participate in continual agency debate over program issues, but the debates proceed in a language which means different things to different people. The debates are not centered on a measurable set of program descriptions nor are the assumptions guiding the program intervention made clear enough to be testable. In fact, most of the people in this world will go to great lengths to keep these two things ambiguous in order to expand their area for maneuver. The administrator is a decision maker—he does take action. As in "textbook" management, many of his actions are based on guesses about what is needed, shifting academic opinions and political support, and the demands of a set of higher level policy makers subject to continual turnover. Unlike the "textbook" management, however, the typical government administrator does not establish and test assumptions linking intervention activities to program performance. Typical government administration might be called "pseudo-management," because all its management activity takes place in a process that is not linked to actual program results. In its own terms, such "pseudo-management" is good if its activities remain acceptable to an ever-changing cast of characters at the policy level.

Evaluators and pseudo-managers operate independently of one another. There is no basis for communication between them. The pseudo-manager has no real use for evaluation and the evaluator can provide few, if any, services to assist in pseudo-

management. In fact, sound evaluation results may present a clear and present danger to the pseudo-manager. In this environment, the evaluator can expect his work to have minimal impact. The problem for the evaluator is to distinguish pseudo-management from textbook management. On the surface it appears to us that pseudo-management predominates in social agencies; the potential for textbook management is yet unknown.

Our emphasis on identifying actual users of evaluation and on pre-specifying the decision context and uses of evaluation information may seem excessive. Yet the desired use of evaluation information determines not only how much it is worth but also the form and accuracy that it must have. And if those in charge of a program have no use for information about that program, then there is no real way to design an adequate evaluation for them. What this might mean may be demonstrated by an example.

Assume that a federal drug treatment program for heroin abusers defines outcome success in the following terms: the client reveals absence of heroin use six months after discharge from treatment, as tested by three randomly spaced urinalyses during the follow-up period and one urinalysis at the end of the six months. Those in charge of the program say that they require information about this outcome to assist in decisions about the following: allocation of technical assistance among drug treatment projects, reallocation of funds among projects, and assignment of headquarters staff to study problems associated with achieving a desirable outcome level. But suppose those in charge are challenged to specify in advance how decisions might vary with the range of possible evaluation findings. For example, will a task force convene for program redesign if national program cure rates average 5 percent, 15 percent, or 50 percent? Will technical assistance be given to projects whose average cure rate falls below 5 percent? Is there technical assistance to give? Can projects be closed down? Will a stated national objective of a 30 percent cure rate be adjusted downward, if the actual average cure rate found is 15 percent? This type of dialogue would permit the evaluator to assess the potential value of evaluation information by identifying plausible and practical uses for it and also permit the evaluator to assess the specific type and accuracy of the information required.

The level of validity and reliability required in measurable data should be an important factor used in analyzing the method of collection, the cost of data collection, and the methods and cost of data analysis

before data collection efforts ever begin. The "conclusiveness" of data only takes on meaning in relation to particular actions the data may suggest. But as we saw earlier, if everything is left ambiguous, no one will know what level of evaluation findings would or should prompt action and therefore what level of validity and reliability are required in the evaluation data. This means that, in our example, drug program evaluations which show cure rates of 2 percent, 5 percent, 20 percent, 50 percent, and 75 percent could all be dismissed by the pseudo-managers as "inconclusive" for decision making.

When a single individual does not have the authority to take or to elaborate on the kinds of action mentioned above, those individuals whose consensus is required must be found and consulted. The point is that management of a program is a policy matter. Evaluation cannot prescribe management actions. Rather, the needs of management should define evaluations.

THE CONSEQUENCES OF EVALUATING WHEN THESE CONDITIONS EXIST

Why should the evaluator worry about the soft, unmeasurable underbelly of social program goals, objectives, and activity; about the obscure logic of program assumptions; or about whether there is a management vacuum? If our analysis is correct, weaknesses in these areas can disable an evaluation effort while making the failure appear to be the evaluator's own doing.

If the agency evaluator, alone or with a contractor, attempts to carry out an evaluation of a program where these flaws exist, our experience indicates that there are two highly likely outcomes. First, the evaluator's attempt to define the program in measurable and logical terms will flounder. No available methodology can bridge the gap between the program as implemented in the field and the program as suggested by program goal statements. Thus the results of his evaluation are likely to be labeled "inconclusive," "abstract," or "an effort to develop methodology." Second, his findings will not be responsive to the information needs of those in charge of the program. He may produce the wrong information or information that is too imprecise or too sparse. Even if the evaluation is technically unimpeachable, those in charge of the program may find it irrelevant to their decision context, seeing no way to act upon the information.

We have suggested that the definition of measurable program design and of testable assumptions about how the program works is a major policy issue which should

be resolved by policy makers and program managers within the discretionary boundaries of program legislation. Program policy making is not the job of the agency evaluator, and he should not undertake the task even if it is disguised as a "technical" choice of the proper program measures needed to conduct an evaluation.

SOME SOURCES OF LEVERAGE

Is there a strategy that evaluators can adopt to return the jobs of policy making and program management to policy makers and evaluators—and improve the utility and yield of evaluation (and program) dollars? Fortunately, some factors in the present federal environment may supply the leverage needed to force attention to the three conditions (lack of definition, lack of clear logic, and lack of management) that have proven costly to program effectiveness and evaluation.

First, there is less naiveté about federal social programs today. More awareness exists that attacking a vague problem with an unproven social, behavior, or economic theory is not likely to bring success. Raising issues about program definitions and assumptions is now more likely to strike a responsive chord in this climate. Secondly, the federal budget is not expanding rapidly, and the present Administration and the Congress are placing more emphasis on accountability. Third, both the Congress and citizens are pushing for more effective delivery of public services, and more evidence of effectiveness.

The evaluator, with some help from high level policy makers and program managers, may be able to take advantage of these potential sources of leverage and use them to force the definition that makes evaluation possible. At least he may assure that his efforts are expended in areas where there is the best chance of success. The tool that we recommend he employ is a "preassessment of evaluability" for every program that is a candidate for evaluation.

PREASSESSMENT OF EVALUABILITY

We recommend a process of pre-evaluation design.⁴ If conducted in proper detail, this process can provide what might be called a "rapid feedback evaluation" of the present status of a program and its information base, and can make clear whether a major evaluation effort is or is not warranted. In essence, the three root causes of problems in program evaluation can be transformed into a set of criteria for determining the evaluability of a public program. These criteria are expressed in the following questions:

1. Are the problems, intended program interventions, anticipated outcomes, and the expected impact sufficiently well defined as to be measurable?

2. In the assumption linking expenditure to implementation of intervention, intervention to the outcome anticipated, and immediate outcome to the expected impact on the problem, is the logic laid out clearly enough to be tested?

3. Is there anyone clearly in charge of the program? Who? What are the constraints on his ability to act? What range of actions might he reasonably take or consider as a result of various possible evaluation findings about the measures and assumptions discussed above?

In a sense the criteria are sequential. Measurable definitions form a basis for the testable assumptions. Then both serve as a basis for the consideration of the range of decisions that those in charge of the program might make as a result of information about actual costs, interventions, outcomes, and impact.

In practice the evaluator will have to judge the degree to which the three criteria are satisfied for particular programs. The evaluator generally has several programs in his agency that can be evaluated at any one time. In initial planning, the evaluator should focus on testing each program against these three criteria, using the best information available from the programs themselves to assess how valuable each program may be. This assessment should be discussed directly with policy makers and program officials. The interaction between evaluator and program officials may assist policy makers and program officials to define the measures and specify the logic of assumptions that need to be tested.

The next task is to decide which programs meet all three criteria. Then programs that meet some criteria, or almost meet all criteria, may be sorted out. Finally, in most agencies, a third group of programs will emerge which satisfy few—if any—of the three criteria.

At this point the evaluator will have completed his own preassessment of the "evaluability" of the programs of his agency. It is almost useless to explore questions of use and methodology for programs that

clearly do not meet the criteria. The next and final step is both a possible source of leverage for the evaluator and a somewhat risky business in many agencies.

CLEARLY NAMING THE PROBLEM FOR OTHERS

The evaluator has now created three lists of programs: "evaluable," "potentially evaluable with further program or management definition," and "not evaluable." Since these problems are now understood to involve policy and management questions, as well as evaluation design questions, the list has two uses.

First, the evaluator should evaluate only the programs that are evaluable. He should agree to help with the definitional problems of potentially evaluable programs. But he should not hesitate to name the nature of the problem. The evaluator should tell policy makers and program managers whether their programs are or are not evaluable, and why. Second, the evaluator should bring the serious problems on the list to the attention of the top level of the agency hierarchy so they will know which programs are or are not evaluable, and why.

These actions may be very risky things to do in many agencies, but it can prevent a lot of useless evaluation attempts and later recrimination. We believe that they would force improvements in program performance as well.

Notes

1. Joseph S. Wholey et al., *Federal Evaluation Policy* (Washington, D.C.: The Urban Institute, 1970).
2. Garth N. Buchanan and Joseph S. Wholey, "Federal Level Evaluation," in *Evaluation*, vol. 1 no. 1 (Fall 1972), pp. 17-22.
3. For a concise overview of the literature in which these criticisms have been put forward, see Francis G. Caro, ed., *Readings in Evaluation Research* (New York: Russell Sage Foundation, 1971), pp. 9-15. In our own work we have had access to unpublished internal assessments of evaluation efforts by several federal agencies; the majority of these note agency dissatisfaction with their evaluation product and identify many of these apparent causes as major influences.
4. See John D. Waller and John W. Scanlon, *Urban Institute Plan for the Design of an Evaluation* (Washington, D.C.: The Urban Institute, March 1973).

Formulating the Question and Measuring the Answer

Carol H. Weiss

The traditional formulation of the evaluation question is: To what extent is the program succeeding in reaching its goals? Variations are possible: Is program *A* doing better than program *B* in reaching their common goals? How well is the program achieving results *X*, *Y*, and *Z* with groups *F*, *G*, and *H*? Which components of the program (*R*, *S*, or *T*) are having more success? But the basic notion is the same. There are goals; there is a planned activity (or several planned activities) aimed at achieving those goals; there is a measure made of the extent to which the goals are achieved. In evaluation there is also the expectation that controls are set up so that the researcher can tell whether it was the *program* that led to the achievement of goals rather than any outside factors (such as the maturing of the participants, improvement in the economy, and so on)

The evaluation question sounds simple enough in the abstract. All the researcher has to do, it seems, is:

1. Find out the program's goals.
2. Translate the goals into measurable indicators of goal achievement.
3. Collect data on the indicators for those who participated in the program (and for an equivalent control group who did not).
4. Compare the data on participants (and controls) with the goal criteria.

And voilà!

But what looks elementary in theory turns out in practice to be a demanding enterprise. Programs are nowhere near as neat and accommodating as the evaluator expects. Nor are outside circumstances as passive and unimportant as he might like. Whole platoons of unexpected problems spring up. This chapter deals with four:

1. Program goals are often hazy, ambiguous, hard to pin down. Occasionally, the official goals are merely a long list of pious and partly incompatible platitudes.

2. Programs not only move toward official goals. They accomplish other things, sometimes in addition and sometimes instead. The evaluator has a responsibility to take a look at these unexpected consequences of program activities.

3. The program is a congeries of activities, people, and structures. Some of its elements are necessary for the effects it achieves; others are irrelevant baggage. Decision makers want to know what the basic and essential features of the program are, so that (if successful) they can reproduce them or (if unsuccessful) avoid them. How do you identify and separate out the elements that matter?

4. The evaluation question as posed ignores the issue of why the program succeeds or fails. The *why* is often just as important to know as *how well* the program works.

In addressing these issues, we will recommend a series of strategies. Possibly the most important theme . . . is the classification of the component parts of the program. Each element (of activity, approach, structure, participant, and so on) that is presumed likely to affect outcomes is observed, defined, and classified. The differences that evolve between groups, between activities, and so on give increasing information about what works and does not work in reaching program goals.

In this chapter, then, we consider these core issues:

1. Formulating the program goals that the evaluation will use as criteria
2. Choosing among multiple goals

3. Investigating unanticipated consequences
4. Measuring outcomes
5. Specifying what the program is
6. Measuring program inputs and intervening processes
7. Collecting the necessary data

FORMULATING PROGRAM GOALS

It is a common experience for an evaluator to be called in to study the effects of a program and not be told its purpose. If he presses for a statement of goals, program administrators may answer in terms of the number of people they intend to serve, the kinds of service they will offer, the types of staff they will have, and similar information. For program implementers, these are "program goals" in a real and valid sense, but they are not the primary currency in which the evaluator deals. He is interested in the intended *consequences* of the program. When he pursues the question, "What is the program trying to accomplish?" many program people give fuzzy replies, often global and unrealistic in scope. They may hazard the statement that they are trying to "improve education," "enhance the quality of life," "reduce crime," "strengthen democratic processes." Thus begins the long, often painful, process of getting people to state goals in terms that are *clear, specific, and measurable*.

The goal must be clear so that the evaluator knows what to look for. In a classroom program, should he look for evidence of enjoyment of the class? interest in the subject matter? knowledge of the subject matter? use of the subject matter in further problem solving?

The goal has to be specific. It must be able to be translated into operational terms and made visible. Somebody has to *do* something differently when the goal is reached. Thus, if the goal is to interest students in new materials, they are likely to talk more often in class, or raise their hands more often, or do more outside reading on the subject, or tell their parents about it, or any of several other things.

For evaluation purposes, the goal has to be measurable. This is not as serious a restriction as it may seem at first glance. Once goal statements are clear and unambiguous, skilled researchers can measure all manner of things. They can use the whole arsenal of research techniques—observation, content analysis of documents, testing, search of existing records, interviews, questionnaires, sociometric choices, laboratory experiments, game playing, physical examinations, measurement of physical evidence, and so on. With attitude

tests and opinion polls, they can measure even such relatively "soft" goals as improvements in self-esteem or self-reliance. But since few programs set out only to change attitudes, the evaluator will also want to find and measure the behavioral consequences of changed attitudes—the things participants do because they feel different about themselves, other people, or the situation.

Some programs find it extremely difficult to formulate goals in these terms. David Kallen tells of working with an advisory committee to plan for evaluation of a detached worker program for gang youth. Asked to specify the program's goals, the committee members came up with such things as improving the behavior of the youth, helping them become better citizens, and improving their school work. When they tried to translate the goals into operational criteria of program success, "behavior" and "citizenship" were too vague to use, and school grades were too likely to be influenced by teachers' stereotyped perceptions of the youngsters. The discouraging story continues:

Finally, it turned out that a number of the area residents objected to the young people's use of swear words, and it was decided that one measure of behavioral improvement would be the reduction in swearing, and that this was something the detached worker should aim for in his interaction with the youngsters he was working with. [Was the group identifying program goals or making up new ones?] It was therefore agreed that part of the criteria of success would be a reduction in swearing. I might add that this was the only measure of success upon which the evaluation team and the program advisory committee could agree.¹

Fuzziness of program goals is a common enough phenomenon to warrant attention. Part of the explanation probably lies in practitioners' concentration on concrete matters of program functioning and their pragmatic mode of operation. They often have an intuitive rather than an analytic approach to program development. But there is also a sense in which ambiguity serves a useful function: It may mask underlying divergences in intent. Support from many quarters is required to get a program off the ground, and the glittering generalities that pass for goal statements are meant to satisfy a variety of interests and perspectives.

However, when there is little consensus on what a program is trying to do, the staff may be working at cross-purposes. One side benefit of evaluation is to focus attention on the formulation of goals in terms of the specific behaviors that program practitioners aim to achieve. The effort may force disagreements into the open and lead to conflict. But if differences can be

reconciled (and the program may not be viable if they are not), the clarification can hardly help but rationalize program implementation. It may reveal discrepancies between program goals and program content, in which case either the content or, as Berlak notes,² the goal statement should be changed. When a sense of common purpose is reached, the logic and rationality of practice are likely to be enhanced.

What does an evaluator do when he is faced with a program that cannot agree on a statement of specific and meaningful goals? Four courses are open to him:

1. He can pose the question and wait for program personnel to reach a consensus. But as Freeman and Sherwood³ note, he should bring books to the office to read while waiting for them to agree. And they still may not develop a statement that provides an adequate basis for evaluation.

2. Another thing he can do is read everything about the program he can find, talk to practitioners at length, observe the program in operation, and then sit down and frame the statement of goals himself. Sometimes this is a reasonable procedure, but there are two dangers. One is that he may read his own professional preconceptions into the program and subtly shift the goals (and the ensuing study) in the direction of his own interests. The other risk is that when the study is completed, the program practitioners will dismiss the results with the comment, "But that's not really what we were trying to do at all."

3. He can set up a collaborative effort in goal formulation. This is probably the best approach. Sitting with the program people, the evaluator can offer successive approximations of goal statements. The program staff modifies them, and the discussion continues until agreement is reached.

4. He can table the question of goals, and enter not upon evaluation in the traditional sense, but on a more exploratory, open-ended study. In complex and uncharted areas, this may be a better strategy than formulating arbitrary and superficial "goals" in order to get on with the study while the really significant happenings around the program are allowed to take place unstudied, unanalyzed, and unsung. Evaluations based on too-specific goals and indicators of success may be premature in a field in which there is little agreement on what constitutes success.⁴

The experienced evaluator also searches for the hidden agenda, the covert goals of the project that are unlikely to be articulated, but whose achievement sometimes determines success or failure no matter what

else happens. For example, if a program of interdisciplinary studies in a university fails to win the support of the departmental faculties and the university administration, even consummate educational results may not be enough to keep it alive. The evaluator, if he is to study the attainment of goals, is well advised to keep an eye on the "system" goals (those that help maintain the viability of the program in its environment) as well as the "outcome" goals. He will learn much that explains why the program makes the adaptations it does and where the real game is.⁵

Some researchers have even proposed that the goal model of evaluation should be junked in favor of a system model.⁶ The elements of such a model are not yet clear; there are almost as many interpretations as there are participants in the discussion. But the common recognition is that organizations pursue other functions besides the achievement of official goals. They have to acquire resources, coordinate subunits, and adapt to the environment. These preoccupations get entangled with, and set limits to, attainment of program goals. According to system model proponents, an evaluation that ignores them is likely to result in artificial and perhaps misleading conclusions.

What would a system model look like? Etzioni, and Schulberg and Baker suggest that the system model should be based on the evaluator's extensive knowledge of the organization and his understanding of the optimal allocation of resources among organization-maintenance and goal-achievement functions. The key question then becomes: "Under the given conditions, how close does the organizational allocation of resources approach an optimum distribution?"⁷ Provocative as the notion is, it sets such demanding requirements for the evaluator (knowing more about the organization than the organization knows itself) that it is difficult to imagine its practical application, at least in these terms. Perhaps future development will bring its genuine insights into the realm of practicality. For the time being, most evaluators will probably stick with the goal model, which is certainly justifiable on its own grounds, and give as much attention to the organizational and community systems that affect the program as the situation seems to warrant.

CHOICES AMONG GOALS

Once the goals of the project are clearly, specifically, and behaviorally defined, the next step is to decide which of them to evaluate. How does the evaluator make the decision?

Usability and Practicality

Part of the answer lies in the potential for utilization. How will the evaluation findings be applied, and which goals are relevant to that decision? Part of the answer lies in the hard realities of time, money, and access. How far off in time the evaluator can study is limited by how long the project—and the evaluation—lasts; how much he can study is at least partly a function of money; whether he can examine certain classes of effects depends on whether he is permitted access to people and agencies. A tendency endemic in all kinds of research is to study what is easy to study rather than what ought to be studied. It is particularly important for the evaluator to avoid this kind of cop-out and to concentrate on key concerns of the program.

Relative Importance

There remains still another factor—the relative importance of different goals. This requires value judgment, and the program's own priorities are critical. The evaluator will have to press to find out priorities—which goals the staff sees as critical to its mission and which are subsidiary. But since the evaluator is not a mere technician for the translation of a program's stated aims into measurement instruments, he has a responsibility to express his own interpretation of the relative importance of goals. He doesn't want to do an elaborate study on the attainment of minor and innocuous goals, while vital goals go unexplored.⁸

Incompatibilities

In some cases there are incompatibilities among stated goals. A model cities program, for example, seeks to increase coordination among the public and private agencies serving its run-down neighborhood. It also desires innovation, the contrivance of unusual new approaches to services for the poor residents. Clearly, coordination among agencies will be easier around old, established, accepted patterns of service than around new ones. Innovation is likely to weaken coordination, and coordination is likely to dampen the innovating spirit. Which goal is more "real"? Evaluation cannot stick its head in the sand and treat the two goals as equal and independent.

Short-term or Long-term Goals?

Another issue is whether short- or long-term goals are more important. Decision makers, who by professional habit respond to the demands of the budget cycle

rather than the research cycle, usually want quick answers. If they have to make a decision in time for next year's budget, there is little value in inquiring into the durability of effects over 24 months. It is this year's results that count.

But decision makers can often be persuaded to see the utility of continuing an investigation over several years, so that the program's long-term effectiveness becomes manifest. Clearly, it is good to know whether early changes persist, or on the other hand, whether the absence of early change reflects a "sleeper effect," the slow building up of important changes over time. Evaluations, wherever possible, should look into long-term effects, particularly when basic policies or costly facilities are at stake. A comparison of short- and long-term effects provides additional information about how, and at what pace, effects take place.

The evaluator is well-advised to thrash out the final selection of goals for study with decision makers and program managers. They are all involved. It is he who will have to live with the study and they who will have to live with the study results and—one would hope—their implementation.

YARDSTICKS

Once the goals are set, the next question is how much progress toward the goal marks success. Suppose a vocational program enrolls 400, graduates 200, places 100 on jobs, of whom 50 are still working three months later. Is this success? Would 100 be success? 200? 25? Without direction on this issue, interpreters can alibi any set of data. A tiny change is better than no change at all. No change is better than (expected) retrogression. Different people looking at the same data can come up with different conclusions in the tradition of the "fully-only" school of analysis. "Fully 25 percent of the students . . ." boasts the promoter; "only 25 percent of the students . . ." sighs the detractor.

Only on a comparative basis does the question really make sense. How do the results compare with last year's results, with the results for those who did not get the special program, or better still, with the results from programs with similar intent?⁹ If comparable data are not available, the evaluator can present his results and let others draw their own conclusions. Or he can get into the act by drawing on past experience, the opinions of administrators and staff, and perhaps outside experts, in reaching a judgment of his own.¹⁰ Early attention to standards of judgment—before the data come in—can forestall later wrangling.

UNANTICIPATED CONSEQUENCES

The program has desired goals. There is also the possibility that it will have consequences that it did not intend. The discussion of unanticipated results usually carries the gloomy connotation of undesirable results, but there can also be unexpected good results and some that are a mixture of good and bad.

Undesirable effects can come about for a variety of reasons. Sometimes the program is poorly conceived and exacerbates the very conditions it aimed to alleviate. A loan program to inefficient small businessmen may only get them deeper into debt. Or a program can boomerang by bringing to light woes that have long been accepted. Some programs raise people's expectations. If progress is too slow or if only a few people benefit, the results may be widespread frustration and bitterness. Occasionally, a program that invades the territory of existing agencies generates anger, competition, and a bureaucratic wrangle that lowers the effectiveness of services.

Good anticipated consequences are not so usual, because reformers trying to sell a new program are likely to have listed and exhausted all the possible positive results. Nevertheless, there are occasions when a program has a happy spin-off, such as having its successful features taken over by a program in a different field. There can be spillovers of good program results to other aspects of a program participant's life. For example, pupils who learn reading skills may become more cooperative and less disruptive or aggressive in school and at home. Contagion effects appear, too. People who never attended the program learn the new ideas or behaviors through contact with those who did.

Sometimes programs tackle one aspect of a complex problem. Even if they achieve good results in their area, the more important effect may be to throw the original system out of kilter. Thus an assistance program to underdeveloped areas introduces a new strain of rice that increases crop yield—the goal of the program. But at the same time, the effect is to make the rich farmers richer (because they can afford the new seed and fertilizer and can afford to take risks), widen the gulf between them and the subsistence farmers, and lead to social and political unrest. Fragmented programs all too often fail to take into account interrelationships between program efforts and the overall system in which people function. What are originally conceived as good results in one sphere may be dysfunctional in the longer view. It is because of such complex interlinkages that

the notion of a systems approach to evaluation is appealing.

The evaluator has to keep an eye on the "other" consequences of the program he is studying. Although decision makers have not articulated them as goals, he must unearth and study consequences that have significant impact on people and systems. Like the formulation of goals, this exercise requires thought and attention. A wise evaluator brainstorms in advance about all the effects, good, bad, and indifferent, that could flow from the program. Envisioning the worst as well as the best of all possible worlds, he makes plans for keeping tabs on the range of likely outcomes. What were "unanticipated consequences" are now—if he judged well—unintended but anticipated. He also has to remain flexible and open enough to spot the emergence of effects that even his sweeping imagination had not envisioned.

If he or his evaluation staff is close enough to the scene to observe what goes on, informal observation may be sufficient for the first look at unplanned effects. In more remote or complex situations, he will have to develop measures and data-gathering instruments to pull in the requisite information. Once trends become clear and side effects are seen to be a strong possibility, he will want as precise measures as he can devise of what may become the most important elements in the program field. He never wants to be caught saying, "The program (on our outcome measures) was a success, but the patient died."

MEASUREMENT: INDICATORS OF OUTCOMES

After the specific goals have been selected for study, the evaluator's next step is the development of *indicators* to measure the extent to which they are achieved. These indicators of program outcomes are the *dependent* variables of the study.

The evaluator is concerned, too, with the description and measure of other factors. There are the relevant aspects of the program—the inputs—which are the *independent* variables of the study. There may also be *intervening* variables, factors that mediate between inputs and outcomes. These two types of measures are discussed in a later section of this chapter.

Developing Measures

The development of measures, sometimes referred to as "instrumentation," is a demanding phase of the

evaluation. If the evaluator is lucky, and earlier studies have been done in the field or measures have been created that are suitable for the subject of his concern, the task becomes one of locating existing measures. It is worth a fair amount of searching to locate measures that have already proved workable, rather than to create new ones. Much of the trial and error work is done. Also, it is generally possible to find out the responses that earlier investigators derived through the use of these measures, and thus to have available some kinds of comparative data from another population. Comparison helps to pinpoint the characteristics of the evaluation group that are "special" and the extent of their divergence from other groups.

Repeated use of common measures helps to build up a body of knowledge. As different evaluation studies use common indicators of outcome (for example, scores on the same test), it becomes possible to begin to make comparisons about the relative effectiveness of one program against another. (The important qualification is that other factors have to be considered, and preferably be constant, if the judgments are to be fair.) In recent years, several handbooks of tried and tested measures have been published as guides for researchers. The distribution of responses from previous research is often included. See, for example:

Charles M. Bonjean, Richard J. Hill, and S. Dale McLemore, *Sociological Measurement: An Inventory of Scales and Indices* (San Francisco, Calif.: Chandler, 1967).

Oscar Buros, *Mental Measurements Yearbook*, 6th ed. (Highland Park, N.J.: Gryphon Press, 1965).

—, *Personality Tests and Reviews* (Highland Park, N.J.: Gryphon Press, 1970).

Delbert C. Miller, *Handbook of Research Design and Social Measurement* (New York: McKay, 1964).

John P. Robinson, Jerrold G. Rusk, and Kendra B. Head, *Measures of Political Attitudes* (Ann Arbor, Mich.: Survey Research Center, University of Michigan, 1968).

John P. Robinson and Phillip R. Shaver, *Measures of Social Psychological Attitudes* (Ann Arbor, Mich.: Survey Research Center, University of Michigan, 1969).

John P. Robinson, A. Athanasiou, and Kendra B. Head, *Measures of Occupational Attitudes and Occupational Characteristics* (Ann Arbor, Mich.: Survey Research Center, University of Michigan, 1967).

Marvin E. Shaw and Jack M. Wright, *Scales for the Measurement of Attitudes* (New York: McGraw-Hill Book Company, 1967).

US Bureau of the Budget, Executive Office of the President, *Household Survey Manual 1969*. Gives concepts, definitions, and questions used by federal statistical agencies.

Frequently, for all the assiduity of his search, the evaluator finds no measures directly relevant to the subjects he wants to study.¹¹ He may find some that are inferentially related, and if they are easy to collect, long-used, or have known distributions, he may be tempted to make do with them. For example, one evaluator wants to know whether mothers who participated in a parents' program become more permissive in raising their children. Let us say that he finds a tolerance inventory, a personality test that has been normed on groups similar to his and includes items on acceptance of deviant behavior. Should he use the mothers' scores on this test as an indicator of increased permissiveness in child rearing? The main assumptions he has to make are (1) that the quality of "tolerance" that the inventory measures is the same dimension as the one reflected in child-rearing permissiveness and (2) that written answers to the inventory questions are a true reflection of mothers' actual behaviors.

It is usually wiser to stick to the relevant core of the subject under study than to rely on a string of unproved assumptions. Developing new measures can be difficult and time-consuming, but measures that are off-center from the main issue, even when reputable and time-honored, are likely to be of little use at all.

Devising questions, test items, and forms often looks so easy that it comes as a shock to find how many people fail to understand or misinterpret even seemingly simple items. Before embarking on the development of new measures, the investigator should have an acquaintance with considerations of validity and reliability. . . . Careful conceptualization and definition are called for, and questions have to be pre-tested and revised (often several times around) until it is clear that they are bringing in the desired information.

Multiple Measures

Adequate indicators of success in evaluation, like adequate measures of concepts in all social research, usually entail multiple measurement. Each specific measure is an approximation of the outcome in which we are really interested. Say we are concerned with

good driving as the outcome of a course in driver education. Knowledge of traffic rules can be one measure; ratings of driving ability by an observer might be another; records of traffic violations, a third. At best, each is a partial measure encompassing a fraction of the larger concept.¹² On occasion, its link to the real outcome is problematic and tenuous. Moreover, each measure contains a load of irrelevant superfluities, "extra baggage" unrelated to the outcomes under study. By the use of a number of measures, each contributing a different facet of information, we can limit the effect of irrelevancies and develop a more rounded and truer picture of program outcomes.¹³

Multiple measures might be useful, for example, in the evaluation of educational curriculums. Evaluators have usually relied on formal tests of pupil achievement. But test scores are influenced by many things other than the cogency of the curriculum, and the curriculum is intended to produce understandings and applications of knowledge that are only partially susceptible to assessment by tests. Cronbach suggests the need for measures of classroom behavior, attitudes, and the subsequent careers of students.¹⁴

Separate measures can be combined to form one overall measure of program success. This requires some assurance that the different measures are complementary and not repetitions of the same dimension. Further, it requires decisions on the relative importance of the different measures (Do they deserve equal billing?) and on the statistical procedures to represent the relative values of measures that use different scales (How do you combine reading scores and numbers of books borrowed from the library?). There is also the possibility that a composite index masks the upward and downward movement of the separate indicators. Therefore, even when an index is a useful device, the evaluator will want to report changes in the separate measures as well.

Proximate Answers

Sometimes the real changes that a program wants to produce lie far in the future and are not so much "goals" as unanalyzed pious hopes. A management incentive program aims to increase the organization's attraction for executives in the interests of improved long-term managerial effectiveness; an educational program for inner-city children is intended to improve their school performance in order to enhance their social and economic status as adults. It would take years, even decades, to test the program's effectiveness in achieving its long-range expectations. In the interim,

proxy measures have to be used that are germane to more immediate goals and presumably linked to desired ultimate outcomes, for example, length of executive tenure, or children's scores on achievement tests. Unfortunately, there is often little research evidence that the purported relationships hold—for example, that tenure is associated with management effectiveness or that school achievement is directly related to the economic and social advancement of the poor.¹⁵ Nor does existing knowledge always suggest better proximate measures.

The problem affects not only evaluation; it is also central to program design. Programs have to be designed to produce certain short-term changes on the assumption that they are necessary conditions for achieving long-range ends. As in many other aspects, the evaluation inherits the fallibilities of the program. Often the best that evaluation can do, at least under the usual time constraints and in the absence of better knowledge, is accept the program's assumptions and find out how well near-term goals are being met. It is left to further research to explore the relationships between short-term goals and long-term consequences. This solution-by-retreat lacks heroic grandeur, but in many cases it represents the only feasible way to get on with the evaluative job.

But when desired consequences are not so remote as to outlast the evaluation, there are decided advantages in measuring both the short-term and longer-range effects. Programs attempt to set in motion a sequence of events expected to achieve desired goals. As Suchman has noted,¹⁶ if the program is unsuccessful, there are two general categories of reasons. Either it did not activate the "causal process" that would have culminated in the intended goals (this is a failure of program), or it may have set the presumed "causal process" in motion but the process did not "cause" the desired effects (this is a failure of theory). (See figure 1.)

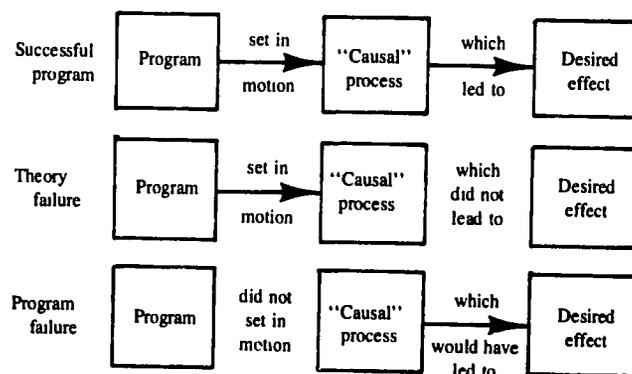


FIGURE 1. Types of program failure.

Stated another way, program failure is a failure to achieve proximate goals; theory failure occurs when the achievement of proximate goals does not lead to final desired outcomes.

When previous research has demonstrated the link between immediate program goals and desired long-term outcomes, there is little need for evaluation to pursue results beyond the first stage. In evaluation of a Smokers' Clinic, it is probably enough to discover that the program led participants to stop smoking. It is not essential to investigate the ultimate incidence of lung cancer. But in fields where knowledge is less developed, further investigation is highly desirable.

Knowing only that intended effects were not achieved is not instructive for future program planning. The availability of short-term and long-term measures would help to indicate whether this was a result of program failure or theory failure.¹⁷ By examining the association between immediate effects and long-range consequences, evaluation can contribute to program practice and program planning—and also to the development of social theory.

Types of Measures

Measures that are useful for assessing outcomes in evaluation research depend on program intent. They can deal with attitudes, values, knowledge, behavior, budgetary allocation, agency service patterns, productivity, and many other items. They can relate to the people being served, the agencies offering service of those affected by changed patterns of service, the neighborhood or the community, or the public at large.

Measuring effects on persons served. Most evaluations tend to concentrate on changes in program participants. They commonly use measures of attitudes, values, personality variables, knowledge, and skills. Each of these may be directly relevant to program goals. A vocational education program, for example, may expect to impart basic knowledge, specific skills, values about work, and certain attitudes toward the job. The usual sources of data are tests, questionnaires, and interviews.

Most programs intend to change overt behavior as well. Information on behavior can be collected through self-reports of the individuals in the program or reports by people who know them (teachers, parents, employers). In some settings, behavior can be observed. Records are another source of information for such items as school grades, arrests, earnings, or hospital admissions.

In the past, evaluators tended to rely heavily on the attitudinal and knowledge measures that were easy to obtain and sidestep the collection of the more difficult behavioral data. But the real payoff for programs is usually a change in behavior. Since it is dubious that changes in attitude or knowledge are necessary and sufficient conditions for behavioral change, the evaluator is well advised to proceed into the behavioral realm.¹⁸ It may be a matter of interest to investigate the extent to which attitudinal or knowledge change is an intervening variable, that is, a link to the dependent variable of behavioral change (see figure 2).

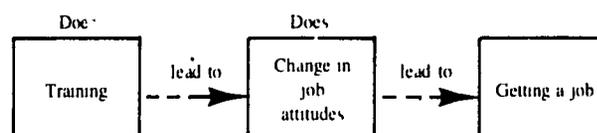


FIGURE 2. Measuring behavioral change—a hypothesis in diagram form

All too frequently, behavioral outcome measures are not directly available. How, for example, do you observe, record, and quantify a decrease in neurotic behavior? The measurement problems are staggering. In many such cases, the evaluator turns to expert judgments. Experts can rate the patient on a scale from “very much improved” to “very much worse.”

When judgments are being used as indicators of outcome, a number of precautions are called for. First, it is usually advisable to have outsiders, rather than the program staff, do the rating. However objective they are, staff members can be suspected of bias—often justly—in the direction of seeing improvement where none exists, or placing high value on tiny, subtle shifts that seem trivial to others. Another precaution is the use of several judges, each rating the same cases, to see how much consistency (reliability) there is between judgments. If interrater reliability is low, the whole procedure is suspect and should be revised. Finally, raters need clear direction, and perhaps training, in the factors to consider and their relative importance. Only when they are each applying the same yardstick, and doing so in terms the evaluator has defined, will their ratings have much meaning.¹⁹

Participants' opinions about the program are sometimes used as an “outcome” measure. They are asked whether they liked it, whether it helped them, whether they would recommend it to others, and similar questions. There is some merit in finding out whether the program appealed to its audience. In the extreme case, if it is totally unacceptable, nobody will come. But

certainly people may like or dislike a program for reasons unconnected with its goals. They may have hazy or misguided expectations of what the program is intended to accomplish and therefore assess its interest and utility in irrelevant terms. Unless it is a specific goal of the program to interest or entertain or offer expected services to the participants, the popularity contest model for evaluation is woefully incomplete.²⁰

Measuring effects on agencies. Some programs aim to produce changes in institutions rather than in people. They may seek to make local bureaucracies more responsive to residents, alter the type of clientele served or the composition of governing boards, broaden an agency's concept of its mission. In this case, indicators of program outcome will be measures of institutional characteristics. Some of these can be aggregated from data about individuals in the organization (for example, percentage with daily contact with local residents), and some can be global measures of the institution (for example, total budget, proportion of the budget devoted to a particular activity, hours during which facilities are open). A useful guide to measures on institutions is Allen H. Barton, *Organizational Measurement*.²¹

In some fields, such as education and public health, there has been a tradition of using "checklist" items developed by experts as "standards of service" as the criterion measures for evaluation. These are generally not outcome measures, but statements of popularly accepted "good practice" (teacher-student ratio, adequacy of equipment). Such measures have been useful for purposes of monitoring an agency's activities, for accreditation, and for educating staff and public about service criteria, but they are not the measures with which evaluation research is primarily concerned, since they relate to program input rather than to outcome. As we will see in the next section, measures of this kind may be useful as intervening variables, indices of particular program features that are presumed to have a beneficial effect on outcomes, to mediate between the program and its effects. Evaluation provides an opportunity to *test* whether supposed "good-practice" components of a program *are* in fact related to successful outcomes.

Measuring effects on larger systems. There are occasional programs whose goals are to make changes in a whole network of agencies (increase communication and referral among all agencies dealing with troubled youth in a community) or to change a community or even a national service delivery system (education, mental health, job training). In such cases, inventive

measures must be devised. Data can come from interviews with leaders; observations within departments or at meetings; collected statistics on clientele, budgets, housing starts; logs kept by staff; analysis of documents or of news stories; sample surveys of the public; and so on. There has not been much evaluation at this "macro" level, but the possibilities are intriguing.

Measuring effects on the public. If a program seeks to alter public values or attitudes, the appropriate indicator of outcome is obviously the public's views. Elaine and John Cumming, for example, administered questionnaires and conducted interviews to determine the effects of a community educational program designed to alter public views about mental illness.²²

SPECIFICATION OF THE PROGRAM

Just as important as conceptualizing the desired outcomes is conceptualizing the nature of the program. What is this program that the evaluator is studying—this amalgam of dreams and personalities, rooms and theories, paper clips and organizational structure, clients and activities, budgets and photocopies and great intentions?

Social programs are complex undertakings. Social program evaluators look with something akin to jealousy at evaluators in agriculture who evaluate a new strain of wheat or evaluators in medicine who evaluate the effects of a new drug. These are physical things you can see, touch, and—above all—replicate. The same stimulus can be produced again, and other researchers can study its consequences—under the same or different conditions, with similar or different subjects, but with some assurance that they are looking at the effects of the same *thing*.

Social programs are not nearly so specific. They incorporate a range of components, styles, people, and procedures. It becomes difficult to describe what "the program" really is. In some cases, a program can be expressed in terms that appear clear and reproducible; for example, a new mathematics curriculum, a change in highway speed limits, a decrease in the size of probation officers' caseloads. But the content of the program, what actually goes on, is much harder to describe. There are often marked internal variations in operation from day to day and from staff member to staff member. When you consider a program as large and amorphous as the poverty program or the model cities program, it takes a major effort to just describe and analyze the program input.

Why should the evaluator be concerned with program input? Haven't we noted earlier that his job is to

find out whether the program (whatever it is) is achieving its goals? Does it make any difference to his work whether the program is using rote drill, psychoanalysis, or black magic? There are evaluators who are sympathetic to such an approach. They see the program as a "black box," the contents of which do not concern them; they are charged with discovering effects. But if the evaluator has no idea of what the program really is, he may fail to ask the right questions. Perhaps because he believes the inflated barrage of program propaganda, he expects mountain-moving outcomes from what are really puny efforts. More likely, he looks for the wrong order of effects. He looks for the attainment of the types of goals that have been verbalized, when the main resources of the operating program have been invested in a different course of action. (If it aimed to teach blacks and ends up directing its attention at whites, is this a course of a different color?)

Furthermore, unless there is some reasonably accurate and coherent definition of the program, the evaluator does not know to what to attribute the outcomes he observes.²³ Let's remember that evaluation is designed to help with decision making. Decision makers need to know what it was that worked or didn't work, what it is that should be adopted throughout the system or modified. Unless the evaluation can provide evidence on the nature of the program as it existed (not merely on the program as described by practitioners), there is little basis for decision.²⁴ In an extreme case, when a program is a smashing success and forty communities want to adopt it forthwith, what is it that we tell them to adopt?

The evaluator has to discover the reality of the program rather than its illusion. If he accepts the description given in the application for funds or in publicity releases, he may be evaluating a phantom. He will be attributing effects (or "no effects") to a program that never took place at all, or to one that operated at so low a level of competence or in such a different manner that it hardly deserves to be called by the program name. For example, teachers who are supposed to be using new curriculum materials may not have received them or may not be following the instructions for their use. A recreation program for school children may be closed on weekends when children have free time; it may be closed erratically so that the children get discouraged from attending; it may be offering activities that fail to attract participants. It will hardly be necessary to collect data on the "effects" of these programs, because there is little or no input. Similarly, Hyman and Wright cite the example of a rural health program in Egypt. The

investigators checking on the staffing of the program found that most of the health centers lacked vital categories of personnel, and that even the people who were employed were putting in relatively few man-hours of work. Thus they could hardly be providing the kind and scope of services intended.²⁵

It takes collection of systematic information and/or observation to find out what is actually taking place. If the program stimulus is well understood and coherent, a few rough measures are sufficient. TB X-rays are TB X-rays, and it is probably enough to know where and to whom they are offered. But if the program is vague or novel or being developed as it goes along, the evaluator may need to describe what is going on. A community mental health center is offering "mental health counseling" to clients, but what exactly does this mean? It may take some digging to find out whether this is referral to psychiatric services, psychotherapy of one type or another, moralizing sermons, referring the client to a job, or any of a dozen other things. Procedures for monitoring the program have to be established. These can be as simple as a discussion with the director, a review of staff records, or attendance at staff meetings, or they can entail interviews with staff or even frequent observation of the program in process. With appropriate sampling procedures, we can characterize even programs that extend widely in space and time.

In most cases, a few relevant descriptive categories suffice to capture the essence of the program. These might include the type of service given, its conceptual emphasis, the type of staff, the setting, and the organizational auspices. If the program is exceedingly complex or differs strikingly from expected patterns, more detail is warranted. Hyman and Wright warn against spending too much time on overelaborating program description; after an initial check to see that the program is really happening and a basic conceptualization, they suggest that the evaluator get on with the job of studying outcomes.²⁶ But circumstances differ. One program's overelaboration is another program's clarification. More precision is obviously called for when program inputs vary. Some clients get one type of service, and others get something different. Under these circumstances, program variables (variables are by definition things that vary) require further attention.

MEASUREMENT: INPUT VARIABLES AND INTERVENING VARIABLES

Not everyone encounters the same experiences within a program. Just as there are outcome variables,

so there are program variables. Some participants in a group therapy program attend every session; others irregularly. Some receive the attention of experienced therapists; others, of relative novices. There are same-sex groups and both-sex groups, groups in the hospital and groups in the community, psychoanalytically-oriented groups and groups of other persuasions.

It is important to look at program variations for two reasons: (1) They clarify the meaning of "the program." They fill in the details of what the general program description has outlined; they show the range of elements that are encompassed by the program—that is. (2) They contribute to the analysis of which features of the program work and which do not. It becomes possible to look at the effects of program components and see whether some are associated with better outcomes than others. If people who are exposed to one kind of service do better than those who are not, we learn something about the relative effectiveness of different strategies and get a clue as to *why* the program is working. If some kinds of people benefit and others do not, we have further notions about the process of change. The analysis of program variables begins to explain why the program has the effects it does. When we know which aspects of the program are associated with more or less success, we have a basis for recommendations for future modifications.

Input Variables

Thus, it becomes important to define and quantify the input variables. They may have to do with variations in:

1. purpose
2. principles
3. methods
4. staffing
5. persons served
6. length of service
7. location
8. size of program
9. auspices
10. management

If the program participants are seen as "put in" to the program, characteristics of participants can also be classified as input variables. Such participant measures can include:

1. age
2. sex
3. socioeconomic status
4. race
5. length of residence in community

6. attitudes toward the program (or toward health, job, marriage, movies, leisure, or whatever else is relevant)

7. motivations for participation
8. aspirations
9. expectations from the program
10. attitudes of other family members toward the program (toward health, job, marriage, and so on)
11. degree of support from family (friends, fellow workers, supervisors, and so on) for the intended goals of the program

Some of these are characteristics that the program cannot affect (age, sex), but others may change as the program goes on (aspirations, peer support). It can be important to have data on both kinds of characteristics to help define who it is that the program helps or does not help.

As evaluations of larger scope are undertaken and as national and international programs are studied, it becomes increasingly important to measure the variations among the units; that is, to measure input variables of each local Head Start, work-incentive program, concentrated employment program, Peace Corps program, and so on. Only then will it become possible to move away from blanket endorsements ("it works") and wholesale rejections ("it doesn't work") and toward the specification that it works or doesn't work under such-and-such conditions.

A note of caution is in order here. There are many variables that are interesting to study. We will even go on and talk about more. But most evaluations have limited resources, and it is far more productive to focus on a few relevant variables than to go on a wide-ranging fishing expedition.²⁷ Ideally, the determination of which input variables are relevant should be based on prior research, but often clear evidence is lacking. After considering the range of possible variables, the evaluator usually has to make his selection on the basis of scraps of data, the accumulated folk wisdom of practitioners, or the application of theory. Until research provides better information, these are not negligible sources of plausible hypotheses. As a rule, it will be more useful for decision purposes to study the factors the program can change ("manipulable variables," such as type of service given) than to focus on fixed attributes over which the program has little control.

Intervening Variables

There can be a further phase in the measurement effort—the specification and measurement of conditions between program inputs and outcomes.²⁸ The

reason for giving systematic attention to these intermediate factors is the expectation that they will affect outcomes. If certain conditions obtain, outcomes will improve; if these conditions are not present, the likelihood of positive outcomes is lessened.

Program-operation variables. There are two kinds of intervening variables. One kind has to do with the implementation of the program—how the program operates. Thus, some participants in a mental health program may have one therapist throughout the program, whereas others (because of staff turnover, shifts in assignments) are served by a succession of different therapists. “Therapist continuity” is a program-operation variable that may influence the ultimate success of treatment.

Other examples of program-operation variables are: frequency of exposure (Do participants who attend 80 percent or more of program sessions do better than those with poorer attendance records?), degree of acceptance by peers (Do group members who are well-liked perform better than those with marginal group status?), extent of coordination of services (Do patients who receive all health services at one health center improve more than those who shuttle around to a number of different health facilities?). Mann offers a list of variables that can affect outcomes in behavior-change programs including the extent of opportunity for practice of new behavior patterns, degree of stress, and amount of participation of the participant and the practitioner.²⁹ Many factors of this sort can have consequences for the degree to which a participant benefits from the program.³⁰

Bridging variables. The other kind of intervening variable has to do with the attainment of intermediate milestones. The theory of the program posits a sequence of events from input to outcome; in order to reach the desired end, certain sub-goals have to be achieved. In a rehabilitation program for prison inmates, it is assumed that learning a skilled trade will reduce a man’s chances of resorting to crime after release. Accordingly, we can measure the extent to which inmates have mastered the skill being taught in the program (as the bridging variable), and then relate skill level to the outcome measure, noncommission of crime.³¹ Other bridging variables might be measured as well, such as length of time between release and getting a job, job earnings, etc., since they reflect underlying assumptions of the program. Changes in participants’ attitudes or knowledge are sometimes hypothesized as necessary preconditions for change in behavior.

Measures of attitude or knowledge change are then viewed as bridging variables.³²

In sum, bridging variables are presumed to link the events of the program to the desired effects, and they represent the theory of the program. On the other hand, program-operation variables may be postulated (or discovered) as necessary conditions for the theory to operate. The two kinds of intervening variables have different implications for the planning of future programs. The bridging variable gives information about the relationship of sub-objectives to final goals. It tests the viability of underlying theory and alerts planners to modifications in assumptions or alternative theories. It might turn out, for example, that the program should be striving to reach different intermediate objectives that would be more effective links to final outcomes. The program-operations variable, on the other hand, contributes to understanding how the program has its effects, what the conditions are for effective operation. Analysis sensitizes planners to features that should be built into the program if it is to operate successfully.

Sources of intervening variables. Intervening variables are usually constructed out of the theoretical assumptions of the program. There are almost always some prevailing notions, however inexplicit, that certain intermediary actions or conditions will bring about the desired outcomes. Intervening variables can also be developed empirically during the course of analysis. Although we may not have had the foresight at the outset to suspect that they were relevant, we can ask questions of the data later, and by appropriate analysis find out whether outcomes are affected by such things as size of group, program cost per pupil, frequency of staff turnover, and so on. The only requirement is that the data must be there to be analyzed—which means the evaluator needs a modicum of cleverness or luck.

Using a model. How do we decide which variables we are going to want to measure? One way is to construct a model of the intended processes of the program. We try to identify the means and the steps by which the program is intended to work. For example, a program of home visits by teachers is inaugurated with the ultimate objective of improving children’s reading achievement. How are home visits expected to improve pupil performance? We might hypothesize the sequence of events shown in figure 3.

The model indicates the kinds of effects that should be investigated. Once ways are found to measure each set of events and the measurements are made, it is possible to see what happens, what works and what

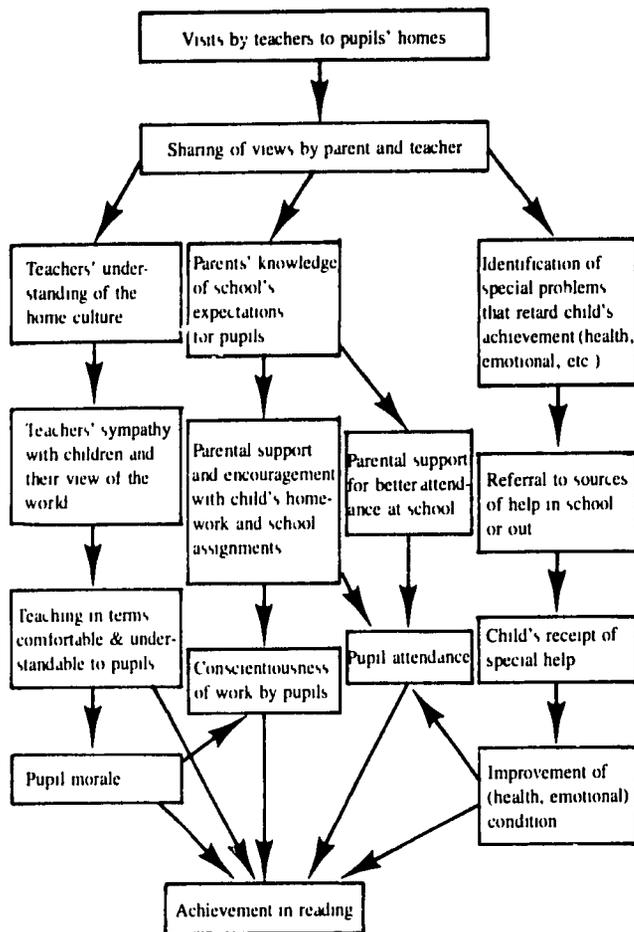


FIGURE 3. A model of a hypothetical program

doesn't, for whom it works and for whom it doesn't.³³ In our home visit program, it may turn out that teachers do indeed show enhanced understanding of the culture of pupils' homes, that they have more sympathy with the children, but that they continue teaching in the old traditional ways without attempting to suit instruction to the pupils' subculture. If the children's reading performance does not improve, we have some clue about the why of it. We can tell where the projected chain has broken down. Similarly, if some parents do learn about the school's expectations for their children's behavior and performance and do try to encourage them to do better homework and schoolwork, yet their children's work is no better done than that of other children, we have a place to look for further insight into the breakdown of the expected chain of events.³⁴

One of the side advantages of setting down the expected paths of change is that it sensitizes the evaluator to shifts in program strategy that make his evaluation design irrelevant. Suppose the home visit program has had difficulties in operation, and to overcome them the

managers have shifted course. They have found that parents really want to talk about things other than home culture and school expectations. To maintain the parents' interest and cooperation, they have shifted to an emphasis on answering parental requests for information (about housing, health facilities, and so on). The original process model is obsolete and must be modified to reflect the new conditions. Some new processes are hypothesized and some old ones discarded; new items and directional lines are added and old ones dropped. The model is adapted to current realities, and measurements and analyses follow the new course. The evaluator, alerted early to the changes in program operations and assumptions, keeps his study flexible enough to be relevant.

The mere construction of such a model can be a useful exercise for program developers. Some of the assumptions that are implicit in the program are made explicit, and naive and simplistic expectations are subject to scrutiny. Do program developers, for example, believe that those parents who do not have the values, the background experience, or the skills to help their children with academic studies can be persuaded to do so by one or two visits by a teacher? Certainly, more must be built into a program with such great expectations—training for the teachers, for one thing, and reinforcements, rewards, and possibly skill training for the parents. The evaluation model can be a learning tool long before the evaluation begins, if program people will use it as such.

A model is not the only way to go about the delineation of necessary measures, but it is one way to clarify and systematize the factors that are worth examining. Suppose that after home visits, pupil reading achievement improves significantly more than that of the control group students who did not have the visits. The usual conclusion would be that the visits (and all the foggy assumptions and expectations surrounding them) were justified. But suppose also that records of teachers' visits showed that all the measures of interaction, communication, and understanding between parent and teacher were at very low levels—the teachers and parents were really not communicating. The source of the children's improved reading ability must be sought elsewhere. (One explanation might be that the students misperceived the intent of the teachers' visits; they may have thought teachers were complaining about their work and trying to persuade parents to punish poor achievers. Improved school work was a response to the perceived threat.)

If the predicted sequence of events does not work out, further investigation is needed. But when the model proves to be a reasonable picture of how things happen, it gives some notion of the reasons why and may be worth further testing. Even with the best and most supportive data, of course, models are never "proved." At best, they are not "disconfirmed" by the data. There may be alternative models that would provide equally plausible or better interpretations of the available facts. Scientific generalizations are built up by developing hypotheses and then submitting them to successive tests in an effort to disprove them or find the limits of their applicability.

COLLECTION OF DATA

Data for evaluation research can come from a gamut of sources and be collected by the whole arsenal of research techniques. The only limits are the ingenuity and imagination of the researcher. Some possible sources are:

1. interviews
2. questionnaires
3. observation
4. ratings (by peers, staff, experts)
5. psychometric tests of attitudes, values, personality, preferences, norms, beliefs
6. institutional records
7. government statistics
8. tests of information, interpretation, skills, application of knowledge
9. projective tests
10. situational tests presenting the respondent with simulated life situations
11. diary records
12. physical evidence
13. clinical examinations
14. financial records
15. documents (minutes of board meetings, newspaper accounts of policy actions, transcripts of trials)

Asking People

Much evaluation research relies on interviews and questionnaires to collect information about program participants—who they are, what they do in the program, and what their attitudes and behaviors are before and after program participation. Staff are frequently queried, too. Tests are a staple ingredient in the evaluation of educational programs. They provide important data on knowledge and learning. Ratings by experts are common in studies of social work, medical, and psychiatric programs.

Observation

On occasion, investigators find ways of collecting relevant data by "unobtrusive" methods that do not involve *asking* anybody anything. Thus, attendance at a science exhibit can be gauged by the number of turns of the turnstile; popularity of exhibits can be measured by the wear in the floor tiles in front of the cases or the number of requests for sample kits. People's adherence to pedestrian safety rules before and after a safety program can be measured by counting the number of people who cross against the light or outside the crosswalks. In health programs, clinical examinations and diagnostic tests have been used. For imaginative examples of these kinds of measures, take a look at Webb et al., *Unobtrusive Measures*.³⁵

Observation is an important tool for collecting data on both pre- and post-program indicators and on the process itself. For maximum reliability, observations should be recorded immediately; if they lend themselves to easy classification, they can be coded on the spot.

Program Records

Program records and agency files are a "natural" for evaluation data. Programs usually collect a fair amount of information about the people who participate. Whether they are pupils in a school, residents in public housing, recipients of welfare, or patients in a hospital, participants will have filled out lengthy forms giving all kinds of information about themselves and their situation. Unfortunately, experience has shown that organizational records are nowhere as useful as they should be. The organization's record-keeping, the transfer of intake and service information to permanent records, tends to be haphazard. Records are inaccurate, out of date, months behind on entries. Furthermore, the definitions and categories used by the agency may be inappropriate for evaluation purposes. Vital categories of information may never have been requested, or records may be kept in a form (such as narrative case-recording) that is inordinately difficult to reduce to items usable for research.

Incompleteness plagues many agency systems. If the participants do not supply certain items of information or if the staff fails to enter data, nobody checks on the missing items and follows up. Thus, if an evaluator needs to know which college students hold outside jobs, he cannot rely on the data in the files for a complete count. Agencies sometimes change record-keeping procedures. If this happens during the period under study, it can vitiate all attempts at before-after

comparisons. Just to add another horror to the gallery, there is the possibility of distortion. Agency records are often based on the reporting of practitioners, and when they know that they are being "judged" by the data in the records, they may intentionally or unintentionally bias their accounts.

On the other hand, there are compensations in the use of institutional records. One is the saving of the time and money that original data collection requires. Another is the advantage in continuity. Unlike the one-shot evaluation study that collects elegant information for a short period of time and then closes up shop, the agency reporting system can provide continual feed-in of information. As our caveats have hinted, few systems will be usable as is. The evaluator will usually have to revamp procedures, introduce new items suited to evaluation requirements, and institute checks for accuracy and completeness. If this can be done and maintained, indicators of program success are constantly on tap for on-going evaluation.

A bail bond project effectively used simple record-keeping procedures to demonstrate its worth. The program involved defendants who could not afford bail before trial. The program staff investigated to see if they had a family, home, job, or other roots in the community, and if they did, arranged to have them released from jail without bail prior to trial. In the three years that the Vera Institute of Justice ran the experiment, 3,505 persons were released. Researchers evaluated the project by looking at court records of appearance (and found that only 56 failed to appear in court).³⁶

For too long, agency record-keeping was relegated to a back office and a superannuated clerk. With the coming of age of computers and the change in perspective from "file cards" to data banks and information systems, institutional records have a better chance of gaining top-level attention and being upgraded. But once the new system is installed and the systems specialists have gone away, there remain a number of vital requirements if evaluation is to benefit. One is that the agency retain an evaluation capacity. Someone has to be able to analyze the products of the information system with an evaluation perspective as well as for descriptive purposes. Another is that the evaluator be in a position to bring the interpreted data to the attention of policy makers, particularly when decisions are pending for which the data are relevant. A further need is the opportunity for periodic revision of the content of the information system. Programs change; today's emphases may be tomorrow's irrelevancies. An ossified information system quickly loses relevance to the live

issues. There has to be a routine way to adapt data items and their presentation to the current scene, without at the same time disrupting items in time series that make for continuity and comparability over time. For some period, old and new items should exist side by side.

Government Records

Some government agencies maintain records on individuals that would be ideal grist for the evaluator's mill. A prime example is the Social Security Administration which collects not only payroll deductions but data—on the number of quarters a person is employed, industry of employment, amount of earnings (up to the Social Security ceiling), and similar information. These data are confidential, but the Administration on occasion will release tabulations of the records of group of people for bona fide research purposes in ways that prevent identification of individuals.³⁷ For evaluation of educational or job training programs, these are extremely useful follow-up data, even when gaps (such as occupations not covered by Social Security) are recognized.

School data are another good source. Depending on school regulations, information on achievement scores, attendance, promotions, and similar items can be retrieved for groups of students. Similarly, there are court records, license bureau records, motor vehicle records, and many other sources of data that are relevant for particular purposes. Automated systems are making data retrieval easier, but it is important to recognize that there are—and should be—limits on the use of records, even for legitimate research. Incursions on people's privacy is a matter of growing concern, and researchers who seek access to official records have a responsibility to limit their requests to essential items and, wherever possible, to use aggregate data (which do not identify individuals) as indicators of outcome.

Government Statistical Series

Another source of data for evaluation is statistical reports of other (usually governmental) agencies. Usually these data have to do with the problem that the program is supposed to "cure." Thus, if a program is designed to reduce illiteracy or highway accidents or infant mortality, the evaluator can turn to statistical reports on the prevalence of these ills and look at changes in the statistics from a time before the program starts periodically through to some later period. If the program is effective, presumably it will tend to push the figures down. Of course, it is necessary that the program and the indicators be coterminous, that they cover

the same geographic area. If the program is working to reduce illiteracy in one neighborhood of St. Louis, it is hardly fair to judge its success by studying the statistics on illiteracy for the whole city.

Scope. Evaluation on the basis of changes in existing indicators is still more a dream than a reality. At the present time there are only limited social data available, and these not necessarily on the most important issues. Much of the data is collected locally, and definitions and categories—even meanings—vary from city to city. Very few data are available for small areas, such as the neighborhood, at which most programs still work.

Accuracy. Furthermore, accuracy is a sometime thing. Crime statistics, for example, have been found to be riddled with reporting peculiarities. There have been at least two newsworthy instances when a new chief of police took office, improved the record-keeping system—and was confronted with a big jump in the crime rate. Victimization studies, surveys that ask representative samples of the public about their experiences as victims of crime, show that large numbers of crimes are never reported to the police. Many statistical series reflect administrative actions as well as the “pure” incidence of the problem. Delinquency rates reflect the activity of the police in apprehending juveniles as well as the actual rate of juvenile crime. Rates of the incidence of diseases depend on the health department’s case finding as well as on the existence of the cases.

Definition of terms. Definitions of terms may also fail to accord with the evaluator’s needs. Standard unemployment figures do not include the underemployed, people who are working part-time but want full-time employment, nor do they count people out of work who have gotten discouraged and are no longer actively looking for a job. Recent Labor Department surveys have found that in poverty neighborhoods, unemployment rates can be substantially understated because of these omissions. If an evaluator wants to use these figures to judge the success of a program for reducing hard-core unemployment, neither the before nor the after figures are suited for his purpose.

Separating effects. In recent years, there has been a movement to develop a national system of social indicators. The proposed system would solve many of the current problems—increase the scope of data, institute common definitions and categories so that data can be compared from place to place, improve accuracy, and certainly allow for disaggregation of data on a small-area basis. There are persuasive reasons for moving toward better social accounting. But as Sheldon and

Freeman note,³⁸ even much better social indicators will not solve all the evaluator’s problems. For evaluation purposes, there is no way of separating out the effects of the program from all the other factors operating on the indicators. If the illiteracy rate goes down, it is not necessarily true that the program was solely, or even mainly, responsible.

Geographic base. Another problem is that indicators are generally based on geography rather than on people. They include any people who are living in the area at a specific time. In some areas, particularly in urban poverty neighborhoods, residents are highly mobile. The people who lived there and were served in the program last year are no longer around to show up in this year’s indicators. It is a whole new batch of people whose status is being measured. If higher-income people have moved into the neighborhood, the changes in the indicators may look exceedingly favorable, even though the real target group is not better off—only farther off.

The fact that indicators cover areas, and thus populations larger than the service scope of the program, minimizes a program’s capacity to make a dent in the numbers. A city will have figures on the percentage of people who vote city-wide and in each precinct, but not figures on voting for the people who were exposed to a get-out-the-vote campaign. It would require changes of heroic proportions in the exposed group to shift the city-wide or precinct percentages of voters.

Inexactness. There is a temptation in using indicators to make do with the figures that exist, even if they are not direct measures of program goals. The use of surrogate measures, or what Etzioni calls “fractional measurement,” is common in all social research, since one indicator rarely captures the entire concept in which we are interested.³⁹ It is probably particularly likely in evaluation-by-social-indicator because of the limited supply of appropriate indicators. For example, we may have a program to improve the quality of housing in an urban neighborhood. There are no available figures on housing quality, but there is a figure on overcrowding, that is, the number of persons per room. The evaluator makes a series of assumptions leading to the conclusion that overcrowding is a reasonable indicator of housing quality, and then draws conclusions about the success of the program on the basis of a measure that at best is only a partial indicator of the true objective.

Manipulation. Indicators are also susceptible to manipulation. Once national indicators are established, program personnel—like the teacher who teaches to the test—may work to improve those facets of their operation that they know will show up and be judged, and pay

less attention to changing the complex social conditions that indicators only partially reflect.⁴⁰

Expectations. Perhaps the gravest impediment to the use of social indicators for evaluation is that it expects so much. A program must be *pervasive* enough to reach a significant part of the relevant population and *effective* enough to bring about change sufficient to shift people from one category to another. A little bit of change is not enough; people have to move from "hospitalized" to "not hospitalized," from "below grade level" to "on grade level," from "unemployed" to "employed." This is asking for program success of giant magnitude. Programs generally reach relatively small numbers of participants and make small improvements. Even the poverty program, considered to be a massive undertaking at the time, was able to mobilize resources that were scanty in comparison with the size of the program. It is little wonder that indicators resist dramatic change.

Even if change does come, it is apt to take a while to show up. Indicators are sluggish. They are derived from periodic soundings, usually annual, so that there is a considerable time lapse before trends become apparent. By the time changes appear in the figures, numbers of other influences have been operating on conditions, and we are back to the problem of authenticating the program as the source of effects.

Some reasonable uses. Nevertheless, there are conditions under which the use of indicators for evaluation would make eminent sense. For massive programs, such as public education or Medicare, they can provide time-series data on the distribution of resources and outcomes.⁴¹ They would have the advantage for federal decision purposes of using common criteria and collecting comparable data across projects and across time, and if astutely constructed, dealing in issues of relevance to policy makers. They cannot overcome such inherent limitations as the failure to account for external (nonprogram) influences or the absence of information on causes and dynamics of change. But if supplemented by, and related to, specifically evaluative studies on critical issues, their information on nation-wide conditions could be supportive and important.

Notes

1. Personal letter from David J. Kallen, January 10, 1966.

2. Harold Berlak, "Values, Goals, Public Policy and Educational Evaluation," *Review of Educational Research*, 40, no. 2 (1970), 261-78.

3. Howard E. Freeman and Clarence C. Sherwood, "Research in Large-scale Intervention Programs," *Journal of Social Issues*, 21, no. 1 (1965), 11-28.

4. See Cyril S. Belshaw, "Evaluation of Technical Assistance as a Contribution to Development," *International Development Review*, 8 (1966), 2-6, 23, for a situation in which this was the case. He goes on, however, to recommend a theoretical framework and a series of possible criteria of success for technical assistance programs, such as an increase in the range of commodities produced or increased division of labor. He offers an approximation of goal statements that can be progressively modified by other researchers, operators, and scholars.

5. Andrew C. Fleck, Jr., "Evaluation Research Programs in Public Health Practice," *Annals of the New York Academy of Science*, 107, no. 2 (1963), 717-24, recommends that evaluators have intimate knowledge of the organization and its relative emphasis on short-run stability versus long-run survival.

6. See Edward A. Suchman, "Action for What? A Critique of Evaluative Research," in *The Organization, Management, and Tactics of Social Research*, ed. Richard O'Toole (Cambridge, Mass.: Schenkman Publishing Co., 1970); Amitai Etzioni, "Two Approaches to Organizational Analysis: A Critique and a Suggestion," *Administrative Science Quarterly*, 5, no. 2 (1960), 257-78; Herbert C. Schulberg and Frank Baker, "Program Evaluation Models and the Implementation of Research Findings," *American Journal of Public Health*, 58, no. 7 (1968), 1248-55; Perry Levinson, "Evaluation of Social Welfare Programs: Two Research Models," *Welfare in Review*, 4, no. 10 (1966), 5-12; Herbert C. Schulberg, Alan Sheldon, and Frank Baker, "Introduction" in *Program Evaluation in the Health Fields* (New York: Behavioral Publications Inc., 1970).

7. Etzioni, "Two Approaches to Organizational Analysis," p. 262.

8. Robert E. Stake discusses the evaluator's responsibility for evaluating proffered goals in "The Countenance of Educational Evaluation," *Teachers College Record*, 68, no. 7 (1967), 523-40.

9. This of course limits the question rather than settles it. How *much* better must the program be before it is considered a success? Statistically significant differences do not necessarily mean substantive significance. Perhaps cost-benefit analysis brings the wisest question to bear: How much does it cost for each given amount of improvement? Carol H. Weiss, "Planning an Action Project Evaluation," in *Learning in Action*, ed. June L. Shmelzer (Washington, D.C.: Government Printing Office, 1966), pp. 15-16.

10. Stake, "The Countenance of Educational Evaluation," pp. 527, 536-38, suggests comparisons with absolute standards, with other programs, and with the opinions of experts for judgment of success.

11. Hemphill suggests that "pure researchers" develop better measures and better conceptualizations for the evaluator to use. John K. Hemphill, "The Relationship Between Research and Evaluation Studies," in *Educational Evaluation: New Roles, New Means*, 68 Yearbook of the National Society for the Study of Education, ed. Ralph W. Tyler (Chicago: National Society for the Study of Education, 1969), chap. 9.

12. Amitai Etzioni and Edward Lehman, "Some Dangers in 'Valid' Social Measurement," *Annals of the American Academy of Political and Social Science*, vol. 373 (September 1967), 1-15.

13. Donald T. Campbell and Julian C. Stanley, "Experimental and Quasi-experimental Designs for Research on Teaching," in *Handbook of Research on Teaching*, ed. N. L. Gage (Chicago: Rand McNally & Co., 1963), pp. 203-4.

14. Lee J. Cronbach, "Evaluation for Course Improvement," *Teachers College Record*, 64, no. 8 (1963), 672-83.
15. David K. Cohen, "Politics and Research: Evaluation of Social Action Programs in Education," *Review of Educational Research*, 40, no. 2 (1970), 218-19.
16. Edward A. Suchman, "Evaluating Educational Programs: A Symposium," *Urban Review*, 3, no. 4 (1969), 16.
17. There is a body of research on the ineffectiveness of counseling and group therapy programs for delinquents and inmates, some of which attempts to cope with this issue. See, for example, H. J. Meyer, E. F. Borgatta, and W. C. Jones, *Girls at Vocational High* (New York: Russell Sage Foundation, 1965); Walter B. Miller, "The Impact of a 'Total-Community' Delinquency Control Project," *Social Problems*, 10, no. 2 (1962), 168-91; H. Ashley Weeks, *Youthful Offenders at Highfields* (Ann Arbor, Mich.: University of Michigan Press, 1958); Edwin Powers and Helen Witmer, *An Experiment in the Prevention of Delinquency: The Cambridge-Somerville Youth Study* (New York: Columbia University Press, 1951); William McCord and Joan McCord, *Origins of Crime: A New Evaluation of the Cambridge-Somerville Youth Study* (New York: Columbia University Press, 1969); Gene Kassebaum, David Ward, and Daniel Wilner, *Prison Treatment and Parole Survival: An Empirical Assessment* (New York: John Wiley & Sons, Inc., 1971); William C. Berleman and Thomas W. Steinburn, "The Execution and Evaluation of a Delinquency Prevention Program," *Social Problems*, 14, no. 4 (1967), 413-23; Nathan Caplan, "Treatment Intervention and Reciprocal Interaction Effects," *Journal of Social Issues*, 24, no. 1 (1968), 63-88.
18. The reverse point, that measure of gross behavioral change is incomplete without knowledge of intrapsychic variables, is made by Leonard A. Kogan and Ann W. Shyne, "Tender-minded and Tough-minded Approaches in Evaluation Research," *Welfare in Review*, 4, no. 2 (1966), 12-17.
19. See Robert B. McIntyre and Calvin C. Nelson, "Empirical Evaluation of Instructional Materials," *Educational Technology*, 9, no. 2 (1969), 24-27. They recommend that expert judgments about instructional materials be supplemented by field evaluations of the effects of the materials in actual situations. An interesting use of expert judgment in a field where behavioral data are hard to come by is the medical audit of professional performance; see Mildred Morehead, "The Medical Audit as an Operational Tool," *American Journal of Public Health*, 57, no. 9 (1967), 1643-56.
20. Studies of compensatory education programs find one "universal finding": Regardless of the type of program, duration, or actual results, parents are enthusiastic. Edward L. McDill, Mary S. McDill, and J. Timothy Sprehe, *Strategies for Success in Compensatory Education: An Appraisal of Evaluation Research* (Baltimore, Md.: The Johns Hopkins Press, 1969), pp. 43-44. See also Donald T. Campbell, "Reforms as Experiments," *American Psychologist*, 24, no. 4 (1969) on the "grateful testimonial," 426.
21. Princeton, N.J.: College Entrance Examination Board, 1961.
22. Elaine Cumming and John Cumming, *Closed Ranks: Study of Mental Health Education* (Cambridge, Mass.: Harvard University Press, 1957). The six-month program produced virtually no change in the population's attitudes toward mental illness or the mentally ill.
23. Even if the program is clearly described and defined, the evaluator may not be able to make precise statements of causality, but he at least will know where to look for effects and have a running start.
24. Coleman suggests that there is often a difference between inputs as offered and inputs as received. The loss of input between its disbursement by authorities and its receipt by pupils may be an explanatory variable in analyzing the effectiveness of educational programs. James S. Coleman in "Evaluating Educational Programs: A Symposium," *Urban Review*, 3, no. 4 (1969), 6-8.
25. Herbert H. Hyman and Charles Wright, "Evaluating Social Action Programs," in *The Uses of Sociology*, eds. P. F. Lazarsfeld, W. H. Sewell, and H. L. Wilensky (New York: Basic Books, Inc., Publishers, 1967), p. 745.
26. Hyman and Wright, "Evaluating Social Action Programs," pp. 755-56. They caution, too, against studying the influence of program variables that the agency cannot modify. The evaluator's task is to identify the effects of components that are manipulable.
27. See Samuel A. Stouffer, "Some Observations on Study Design," *American Journal of Sociology*, 55, no. 4 (1950), 355-61.
28. See Stake, "The Countenance of Educational Evaluation," for a somewhat different conceptualization. He speaks of antecedents, transactions, and outcomes. His "transactions" have much in common with what we call intervening program variables.
29. John Mann, "The Outcome of Evaluative Research," in *Changing Human Behavior* (New York: Charles Scribner's Sons, 1965), pp. 191-214.
30. Katz, for example, discusses the importance of program variables on Negroes' intellectual performance in desegregated schools—for example, social rejection and isolation, fear of competition, and perception of threat. Irwin Katz, "Review of Evidence Relating to Effects of Desegregation on the Intellectual Performance of Negroes," *American Psychologist*, 19, no. 6 (1964), 381-99. In more general terms, M. C. Wittrock states that it is important to measure the environment of learning and the intellectual and social processes of learners as well as their actual learning achievement. "The Evaluation of Instruction," *Evaluation Comment*, 1, no. 4 (1969), 1-7.
31. Freeman and Sherwood, "Research in Large-scale Intervention Programs," give two other examples. They hypothesized improvement in reading skills as an intermediate variable on the way to reduction in school drop-outs (p. 19), and a decline in feelings of alienation and anomie as a basic step toward reduction in delinquent behavior (p. 15). Similarly, in the evaluation of in-service training I suggested a three-part analysis: Do the trainees learn the training content? If so, do they put their learning into practice? Are the trainees who practice what they have learned more successful on the job? Carol H. Weiss, "Evaluation of In-service Training," in *Targets for In-service Training* (Washington, D.C.: Joint Commission on Correctional Manpower and Training, October 1967), pp. 47-54. The extent to which immediate and intermediate goals can be divorced from ultimate goals is discussed in Edward A. Suchman, "A Model for Research and Evaluation on Rehabilitation," in *Sociology and Rehabilitation*, ed. Marvin B. Sussman (Washington, D.C.: American Sociological Association, 1965), pp. 66-67.
32. For an analogous discussion in terms of "sub-objectives," see the report on a health service program in O. L. Deniston, I. M. Rosenstock, and V. A. Getting, "Evaluation of Program Effectiveness," *Public Health Reports*, 83, no. 4 (1968), 328-30.
33. As the similarity between figure 3 and a path diagram suggests, path analysis is a useful way of estimating the strength of the linkages. When the necessary conditions for path analysis are met, path coefficients will represent the magnitude of effect for each

arrow. See Hubert M. Blalock, Jr., *Theory Construction* (Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1969).

It might be useful to posit some negative chains, too, and through path analysis find out the extent to which negative side effects are taking place

Another analytic device, particularly useful when one step in the chain is totally or largely conditional on the one before, is the Automatic Interaction Detector (AID). Unlike path analysis, this type of "branch" analysis shows whether the program is producing strong effects for subsamples of the group. This is a matter of both theoretical and programmatic interest, because if there are differential effects, special programs can be developed to fit special groups. John A. Sonquist and James N. Morgan, *The Detection of Interaction Effects. A Report on a Computer Program for the Selection of Optimal Combinations of Explanatory Variables* (Ann Arbor, Mich.: University of Michigan Institute for Social Research, Monograph no. 35, 1964); John A. Sonquist, *Multivariate Model Building: The Validation of a Search Strategy* (Ann Arbor, Mich.: University of Michigan Institute for Social Research, 1970).

34. An interesting analogy with the reasoning behind the construction of program process models comes from Myron Tribus, "Physical View of Cloud Seeding," *Science*, 163, no. 3928 (1970), 201-10. In the complicated business of producing rain by seeding

clouds, he suggests that rather than relying only on statistical measures of rainfall to evaluate success, investigators should seek to understand and measure the chain of physical events that follow cloud seeding and do or do not lead to rain.

35. E. J. Webb, D. T. Campbell, R. D. Schwartz, and L. B. Sechrest, *Unobtrusive Measures: Nonreactive Research in the Social Sciences* (Chicago: Rand McNally & Co., 1966).

36. Herbert Sturz, "Experiments in the Criminal Justice System," *Legal Aid Briefcase* (February 1967), pp. 1-5.

37. For example, Bernard Levenson and Mary S. McDill, "Vocational Graduates in Auto Mechanics: A Follow-up Study of Negro and White Youth," *Phylon*, 27, no. 4 (1966), 347-57.

38. Eleanor B. Sheldon and Howard E. Freeman, "Notes on Social Indicators: Promises and Potential," *Policy Sciences*, 1 (1970), 97-111.

39. Etzioni and Lehman, "Some Dangers in 'Valid' Social Measurement," p. 2.

40. Campbell, "Reforms as Experiments," pp. 409-29.

41. Cohen, "Politics and Research," pp. 234-36, proposes a system of social indicators of schools and schooling that would permit analysis of inputs to schools, outcomes of schooling, and "temporal, geographic, political and demographic variation in both categories."

What Can We Actually Get From Program Evaluation?

Joseph S. Wholey

INTRODUCTION

As an analyst and as a public official, I have been interested for some time in the role that quantitative analysis can play in assisting decisions on public programs—in particular, in the role that evaluation of program results can play.

The essence of *program evaluation*, as I use the term today, is the assessment of program *outcome*—what happened that would not have happened in the absence of the program?—and *relative effectiveness* within national programs—what individual local projects or types of projects work best? The purpose of evaluation is to provide objective feedback to program managers and policy makers on the cost and effects of national programs and local projects, to assist effective management and efficient allocation of limited resources.

Evaluation has come into its own over the past few years. There has been rather wide acknowledgement by public officials of the need to evaluate social programs. Federal legislation has called for it, money has been provided, evaluation staffs have been created or strengthened, and some major evaluation studies have been undertaken.

During this time, we have all learned that evaluation is difficult, takes a lot of time to carry out, and can be very expensive. We have discovered that the information generated by evaluation studies is often incomplete, suspect, and unrelated to the problems at hand. We have found bureaucratic and organizational constraints so formidable that today, after investment of significant resources and effort, *not one* federal agency has an overall evaluation system and few programs are able to make any use of the evaluations

produced. On the whole, *federal evaluation efforts have not been cost-effective* in terms of impact on policy or program development.

In this paper, I consider the points of view of decision makers at two levels—*policy makers* concerned with legislative changes and budget levels—and *program managers* at various levels. What can policy makers and what can program managers actually *get* from program evaluation?

EVALUATION FOR POLICY MAKERS: NATIONAL PROGRAM IMPACT EVALUATION

First, what can policy makers get from evaluation? Two major types of evaluation are of interest to policy makers: national program impact evaluation, which may throw light on the effects of a national program, and evaluation of field experiments and demonstration projects, which may throw light on the desirability of new operating programs while there is still time to learn from experience.

Head Start

The Westinghouse-Ohio University evaluation of the national Head Start program is a leading example of program impact evaluation.¹ The Westinghouse study design was not perfect, but the results of the study probably furnished a correct assessment of the impact of the national Head Start program. It revealed that one, two, and three years after children from low-income families had gone through the Head Start program, there was little or no improvement in their cognitive achievement or motivational attitudes (when compared with similar children in the same communities).

The Westinghouse evaluation of the Head Start program therefore produced generally negative findings. The negative findings, however, did not significantly reduce the budget level of Head Start. Powerful constituencies would have fought any reduction in funding for Head Start. Results that seem to have come from the Westinghouse Head Start evaluation are (1) the "hold" placed on the program—increased funding levels would not be sought; (2) the diversion of some Head Start program funds into experimental child development programs, "planned variations," designed to test whether there are better approaches than those that were being used in the national Head Start program; and (3) the reduction of the proportion of Head Start funds now going into *summer* Head Start (the Head Start component with the *least* apparent value).

Manpower Training

Perhaps a more typical outcome of national program impact studies can be seen in the manpower area. The Department of Labor spends \$4 million per year on evaluation of manpower programs. Yet a recent Urban Institute study for the Joint Economic Committee concluded that: "Differences in research design and wide natural variations within programs have led to unreliable cost and effectiveness findings for manpower training programs. . . . The manpower training benefit/cost studies reviewed had methodological limitations which made it impossible to be sure that the true average results of the manpower programs were measured."²

Cost-benefit studies of national impact programs consume a large share of the Labor Department's evaluation resources. But the results of these studies play almost no part in the administration of Labor Department programs. *Even if* reliable and valid data were being generated in the national program impact studies being done, such studies are not appropriate support for the types of decisions actually made within the Labor Department. National program impact evaluation studies circulate from office to office in the Labor Department without being acted upon or in most cases even read, because Labor Department administrators do not make the types of decisions which these studies are designed to support. There is room for well-designed cost-benefit studies, but not to the exclusion of other, more relevant types of evaluation.

Title I, ESEA

Another area in which we have examined the feasibility and desirability of national program impact

evaluations is that of compensatory education. The federal government spends \$1.5 billion per year on education of disadvantaged children, under Title I of the Elementary and Secondary Education Act. It might seem important to evaluate the national impact of these large expenditures. Yet, a year ago, the Urban Institute urged the Office of Education not to put \$800,000 of its scarce evaluation funds into a national impact evaluation of Title I. The argument concluded:

While Title I program impact evaluation is feasible, it faces severe methodological problems in sample selection, in defining the treatments provided, in designation of comparison groups, and in dealing with student mobility. . . . As a result of both our inability to distinguish Title I services from other services provided to program participants (by local funds) and the fact that Title I accounts for a relatively small proportion of total per pupil expenditures, it will be very difficult to attribute observed changes in student achievement to Title I. . . .

A national program impact evaluation of Title I, even if methodologically feasible, seems undesirable in comparison with more constructive uses of the evaluation funds available. . . . The Office of Education can and should work with those states that are interested to develop better monitoring systems, to improve local evaluation, to locate and document successful compensatory education projects, to distinguish better from worse Title I projects. . . . The Office of Education can also work through State Educational Agencies to improve the usefulness of local project evaluation efforts by subsidizing cooperative local evaluations which utilize at least some common output measures. . . .³

We are not saying that national program studies are never appropriate. What we *do* suggest is that this type of evaluation is too often done for people who need other types of information to help them select among the options in their decision space. Evaluations have been too willing to accept neat, over-simplified decision-making models. While evaluators have always recognized the need to understand thoroughly the programs being evaluated, rarely if ever is evaluation preceded by an analysis of the decision-making process and the constraints on the options open to the decision makers for whom the evaluation is being done. If evaluation results are expected to affect policy or program management decisions, then an analysis of the *planning-management-control* process to be affected and the development of realistic models of this process must become integral parts of evaluation planning and design.

EVALUATION FOR POLICY MAKERS: EXPERIMENTATION

A second type of evaluation important to the policy maker, but less often carried out, is evaluation of demonstration projects and field experiments—areas in

which evaluation is politically and technically more feasible—and may have more chance to influence decisions. The typical demonstration projects demonstrate only that it is possible to spend public funds in a particular way. The results of the “demonstration” usually go unevaluated.

Police Fleet Plan

The Urban Institute’s study of the Indianapolis Police Fleet Plan is an interesting example of an evaluation of a demonstration program.⁴ In the Indianapolis Police Fleet Plan, police patrolmen are allowed to take their police cars home with them for their private use in off-duty hours—thus putting a lot more police cars on the city streets. The Urban Institute worked with the city of Fort Worth, which had some interest in possibly adopting the Police Fleet Plan. While the evaluation results were quite positive in favor of the Police Fleet Plan (auto thefts went down, auto accidents went down, outdoor crime, purse snatching, and robbery went down),⁵ the study’s sponsor (the Fort Worth City Manager) chose not to implement the findings of the study. After the study was published, however, at least one other local government did decide to implement a police fleet plan based on the results of the evaluation study.

Experiments

In the past few years, there has been a new trend in the development of federal programs. Instead of beginning major new programs or demonstration projects designed to be entering wedges for such programs, the federal government has turned toward using *field experiments* in an attempt to find out what is effective and what is not, *before* a program is implemented nationally. True experiments differ from typical demonstration projects in that those responsible exercise control over inputs and process variables—and carefully measure outputs to determine the extent to which the project reaches its objectives. Five years ago, the idea of conducting large-scale social experiments was neither practical nor realistic, for political reasons. The fact that income maintenance experiments are successfully underway and that money has been earmarked for a housing allowance experiment indicates that federal administrators are increasingly willing to take the political risks involved in running a carefully controlled set of experiments.

Outstanding examples of field experiments are OEO’s negative income tax experiment now underway

in New Jersey; HEW’s income maintenance experiments in Gary, Seattle, and Denver; and OEO’s experiments with performance contracting in elementary school education. OEO’s experiments in performance contracting and the proposed experiments with education vouchers are beginning to break some new ground which may prove important in a number of ways. Private sector agencies will be tested and given a chance to develop new educational program models.

There are two ways to introduce the experimental approach into public programs—or two times at which experiments can be introduced: (1) before a major operating program is undertaken; (2) simultaneous with a major operating program. An experimental program may be started as a possible forerunner of a larger social program; or it may be set up to run alongside a large operating program, to learn things that might improve the operating program. (See, for example, the Office of Education’s Follow Through program, in which a dozen or more approaches to the education of disadvantaged children are simultaneously being tested—each in several communities.) There is growing support in Washington for both of these approaches.

EVALUATION FOR PROGRAM MANAGERS

Let’s turn our attention now to evaluation for program managers, those at federal, state, or local level who have responsibility for operating major programs. Over the past two years, members of the evaluation group at the Urban Institute have become more and more convinced that the primary evaluation pay-off (in terms of decisions actually influenced) may be in evaluation that is done in enough detail to get at the effects of operational changes within operating programs. Many program managers really want to know *what works best, under what conditions*. There is a market, a use, for this type of detailed evaluation information.

Following are three examples (or two-and-one-half examples) of evaluation systems designed to help program managers.

Solid Waste Collection

The Urban Institute recently developed a monitoring system for the District of Columbia Sanitation Department.⁶ Inspectors, supplied with reference photographs, drive along city streets and alleyways with a tape recorder microphone in hand. For each block covered, they rate the cleanliness of the block as 1, 2, 3,

or 4 (by comparing the street or alleyway with the reference photographs). This system therefore produces data on the *outputs* of services, not simply inputs or estimates of outputs.⁷ One can imagine this system being used to assess the results of operational changes in Sanitation Department activities (as is now being done in the District of Columbia) or to justify budget requests (once it is determined that particular additional inputs in the way of increased services can in fact produce differences in outputs, for example, moving a neighborhood's streets and alleys from an average condition 3 or 4 [dirty] to a condition 2 [relatively clean]).

Public Schools

Urban public school systems have increasingly been called upon to address and correct major inequities in our society while providing quality education to large, heterogeneous school populations. School personnel are bombarded with numbers, which are supposed to be useful in making decisions affecting the operations of the school system. Rarely, however, are the data which pour out of large school systems relevant to the needs of school system decision makers. If, in the future, school systems are to respond to the challenges they face, then the objectives of education must be clarified and information about the performance of the school system in meeting those objectives must be improved and used effectively.

At present, most local educational evaluation focuses on analysis of special projects that occupy only a small fraction of the input to a particular school, while opportunities are ignored to make comparisons of input and output across the entire school system. Experience has shown that these local project evaluations, usually carried out to fulfill federal requirements, are of little use to local decision makers because their results are neither timely nor comparable. Project evaluations also operate under severe methodological constraints, which often make their results inconclusive.

The Urban Institute is now working with the Atlanta Public School System trying to develop a system for estimating the relative effectiveness of different public schools in the city.⁸ In this project, Atlanta schools are being classified by the economic level of the students (currently measured by proportions of children receiving free lunches or reduced-price lunches) and by the amount of pupil turnover in the school during the year. The Institute is testing the notion that information on the relative effectiveness of schools serving compara-

ble student populations could be useful to the superintendent and his staff. This work is still in the research and development phase.

Legal Services

Some federal agencies are giving attention to improved systems for program monitoring, where evaluation feedback is used directly to assist management decisions (for example, decisions on the refunding of individual projects and decisions on provision of technical assistance or training to projects performing below expectations). The Urban Institute designed, for the OEO Office of Legal Services, a systematic monitoring system that classifies local Legal Services projects into groups according to the kinds of communities in which they are operating, so that projects operating in similar circumstances can be compared with one another.⁹ (Projects are classified by budget size, type of population served, and type of community in which the projects operate.) When feasible, the same evaluators visit the projects within the same class, to enhance the prospects of making valid comparative judgments among projects that are in fact comparable.

The Office of Legal Services monitoring system rates the quality and quantity of the work being done by local Legal Services staff attorneys and provides Office of Legal Services management with estimates of the results achieved by every one of these projects toward Legal Services program goals (to promote economic development, to reform laws and administrative regulations bearing unfairly on the poor, and to provide individual legal services).

Hard work is required to get evaluative information in enough detail and with enough reliability to help program managers—but that's where real pay-offs for evaluation can occur. And this kind of evaluation is also more acceptable to program managers, who after all are the people who have to provide much of the data required for evaluation studies.

The Legal Services monitoring system and the District of Columbia Sanitation monitoring system are alike in their emphasis on outputs. For the Legal Services program, all that could be obtained through the on-site evaluations were relatively soft data on *estimated* outputs. The District of Columbia solid waste collection monitoring system adds the collection of hard output data on the effectiveness of solid waste collection activities—new data not available in any city records.¹⁰

PROBLEMS WITH EVALUATION

Let's turn now to some of the real problems in getting useful evaluation. From the point of view of decision makers, evaluation is a dangerous weapon. They don't want evaluation if it will yield the "wrong" answers about programs in which they are interested.¹¹ On the other hand, decision makers are more advanced in their ability to ask pertinent questions than evaluators are in their ability to provide timely answers at reasonable cost. Valid, reliable evaluation is very hard to perform and can cost a lot of money.¹² Evaluators have real problems in detecting causal connections between inputs and outputs—and in doing so in timely enough fashion to be useful to decision makers. The structure of a program can have an important influence on the technical feasibility of separating the effects of the program from the effects of other, often more powerful, forces *not* under control of the program. To the extent that a program is run as a controlled experiment, for example, the evaluator's chances of separating out causal connections may be greater.

Our reviews have found typical federal program evaluation studies marked by certain design characteristics which severely restrict their reliability and usefulness:

1. They have been one-shot, one-time efforts, when we need *continuous* evaluation of programs.
2. They have been carried out in terms of national programs and are very weak on process data.
3. They have been small sample studies working with gross averages, when we need studies large enough to allow analysis of the wide variations we know exist in costs and performance among projects within programs. These studies have often been accompanied by conclusions and recommendations based on unsupportable or unmeasured assumptions and weak, and often confusing data. In these cases, the evaluation results *should* be ignored by policy makers. Other evaluation studies, while competently conceived, are so severely constrained by time, money, and an inadequate data base that the results at best have only limited significance for policy changes or program improvement.

Experimentation presents new opportunities for the evaluator—and a new set of problems. There are important tensions between the evaluator and the program official, tensions which arise out of the very notion of experimentation. The criteria for selection of sites, the carefully controlled design of the experiments, and the random assignment of participants (or communities) to

treatments are basic to experimental design. The program administrator may not see the utility of such ideas, however. What is so wrong, he may wonder, about calling an existing exemplary program an "experiment"? Or why not choose the people most in need of housing to participate in a housing allowance experiment? The evaluator must woo and win the administrator to the need for preserving the experimental character of the experiment.

Time also presents an enormous problem for the evaluator of experimental programs. As soon as there is sufficient legislative support to fund a series of experiments, there may be enough support to enact such a program nationwide. The concern that legislation will be enacted before the experiments have had time to produce reliable results may lead to pressures for the release of early, less reliable findings. The New Jersey Income Maintenance experiments experienced this pressure. Some early tentative results from the study were released with reluctance and heavy qualifications. If experimentation is to become a major vehicle in policy research, then ways must be found to anticipate and deal with these types of pressures.

Despite the differences, evaluation of experiments has a great deal in common with the evaluation of on-going programs. In both cases, the evaluator must resist the temptation to search for answers to questions that interest *him*, but which may not be high on the list of questions the *decision maker* wants answered. Decision makers will be convinced of the worth of evaluation only if evaluation meets the needs of the decision maker and provides information useful to him.

Conclusion

What can we actually get from program evaluation? From the point of view of a skeptical, but interested policy maker or program manager, evaluation has a mixed record. From the point of view of the analyst, the problems in doing useful evaluation are formidable. Over the past few years, however, there has been some progress, enough to indicate that certain directions in evaluation have promise.

Notes

1. Westinghouse Learning Corporation-Ohio University "The Impact of Head Start: Evaluation of the Effects of Head Start on Children's Cognitive and Affective Development." 1969.
2. Joe Nay et al., *Benefits and Costs of Manpower Training Programs: A Synthesis of Previous Studies with Reservations and Recommendations* (Washington, D.C.: The Urban Institute, 1971).

3. Joseph S. Wholey and Bayla F. White, letter to Dr. John W. Evans, Assistant Commissioner of Education, November 18, 1970. See Joseph S. Wholey et al., *Title I Evaluation and Technical Assistance: Assessment and Prospects* (Washington, D.C.: The Urban Institute, 1971)

4. See Donald M. Fisk, *The Indianapolis Police Fleet Plan: An Example of Program Evaluation for Local Government* (Washington, D.C.: The Urban Institute, 1970).

5. It's worth noting that the Police Fleet Plan study was done primarily using existing effectiveness data on crime rates, accidents, etc., together with development of cost data from city records. It took only a month or two of an analyst's time to put this study together

6. See Louis Blair and Alfred Schwartz, *Improving the Measurement of the Effectiveness of D.C. Solid Waste Collection Activities* (Washington, D.C.: The Urban Institute, 1971).

7. It turned out that the ratings for streets and alleyways were relatively stable within a census tract and over time. Therefore, it was possible to develop an efficient monitoring system using sampling techniques.

8. See Bayla F. White, *Design for a School Rating or Classifica-*

tion System (Washington, D.C.: The Urban Institute, 1970); and Bayla F. White et al., *The Atlanta Urban Institute School Classification Project* (Washington, D.C.: The Urban Institute, 1971).

9. See Hugh G. Duffy et al., *Design of an On-Site Evaluation System for the Office of Legal Services* (Washington, D.C.: The Urban Institute, 1971).

10. The Legal Services monitoring system was developed for less than \$50,000 and is now being used by the OEO Legal Services program to keep track of their 260-odd local projects. The District of Columbia solid waste collection monitoring system was developed for approximately \$70,000 and is now being implemented in the District of Columbia.

11. In some cases, political pressures will simply override the empirical evidence without the formality of a methodological argument. Here, the only recourse open to the evaluator is to publish the results and hope that some other more enlightened or less pressured decision maker with similar problems will make use of the results.

12. The Stanford Research Institute evaluation of the Office of Education Follow Through program, for example, has already cost approximately \$7 million.

Principles and Methods of Program Evaluation

Herbert D. Turner

Program evaluation, as an instrument for management and policy formulation has grown rapidly in both developed and developing country governments and in multilateral organizations. While producing much useful information, this rapid growth inevitably has been accompanied by confusion about purposes, methods, organizational arrangements and, not least, terminology. This paper attempts to summarize a few of the principles, methods and operational lessons learned in program evaluation in the US Agency for International Development the past few eventful years.

In this paper, we define program evaluation as the retrospective analysis of experience to see if we achieved our stated objectives and to determine how and why such objectives were, or were not, achieved. Thus, program evaluation is concerned with results.

It is useful to differentiate program evaluation from routine implementation monitoring and at the same time to note their interdependence. Program evaluation is a discontinuous function. The evaluator is disengaged from day-to-day operations and should be detached emotionally and intellectually from the program or project. The evaluator:

1. examines the relevance of, and need for, the project;
2. questions the design and the underlying assumptions;
3. assesses induced change and progress toward planned targets;
4. identifies unplanned change;
5. attempts to identify causal factors and assess their effects;
6. feeds his findings into redesign and improved execution.

Implementation monitoring is a continuous function. The monitor is intimately engaged in day-to-day operations and is usually emotionally and intellectually involved in the program or project. The monitor is concerned with:

1. the procurement, delivery and installation of resource inputs;
2. adherence to implementation plans;
3. compliance with required standards and procedures.

In projects evaluated by the project management team, the evaluator and monitor may be one, requiring a periodic shift in attitude and behavior during the evaluation.

The two functions are interdependent and necessarily somewhat overlapping. The monitor generates and collects progress data needed by the evaluator; the monitor may call for an evaluation when difficulties arise; the evaluator's findings are translated into replanning actions and implemented by the monitor.

PURPOSES OF EVALUATION

The purposes of program evaluation in most development agencies are several.

1. *As an integral element in project management*, evaluation is concerned with project performance or effectiveness and would include regular periodic evaluations of all ongoing projects on a decentralized basis for purposes of redesign and improved execution. This performance evaluation is intended to deal with three considerations.

- a. Reaffirming the continued relevance of the project in the light of changes over time in host country circumstances.

b. Measuring progress toward planned targets, and where progress is discernable during implementation, at the impact level; examining unplanned results.

c. Determining causality, i.e., what internal elements of project design and what external factors affected project performance and how they operated.

2. *As a device for improving resource allocation and program management*, evaluation would be concerned with impact or significance and would consist largely of centrally managed, highly selective, in-depth ex post studies of the impact of individual projects and sets of projects on development goals. These evaluations would complement the interim impact measurements noted above and would focus on those situations where project impact on sector/program goals was readily apparent only after project termination. The objective is the pinpointing of operationally useful lessons for application elsewhere. These in-depth evaluations differ from the performance evaluations noted above in that they require a greater investment of skills and other resources as well as more extensive and intensive data collection and analysis.

3. *As a means of policy formulation*, evaluation is concerned with the retrospective examination of program issues which are not country-specific. These would include the role of agricultural credit in rural development, the effects of price policies on agricultural production, and the impact of land reform/land tenancy arrangements on rural development, among other issues.

OPERATIONAL PRINCIPLES

Several operational principles have emerged from recent experience which govern the design and evaluation practices of a development organization.

1. The effectiveness of the evaluation process is largely dependent upon the quality, explicitness, and rigor of program and project design; the quality of design is the major limiting constraint in evaluation.

2. Evaluation must comprehend the total program or project. It is neither feasible nor productive to limit the evaluation process to the fractional resource input of a single source or donor.

3. Evaluation should not be conducted for its own sake nor for generating potentially useful information. Rather, evaluation should be decision-driven; that is, evaluation should be undertaken only in response to a need for a decision or for formulation of plans, programs or policies. Consequently, evaluation findings should be packaged and timed for the needs of the decision maker.

4. The responsibility for evaluation should be placed, functionally and organizationally, as close as possible to the user who will base his decision on the evaluation findings, i.e., the feedback loop should be as small as possible.

5. The host country should take the leading role in evaluating donor-assisted projects with the donors playing a supportive role. Where the host country does not have adequate capacity for evaluation, the donors should offer training in evaluation methodology both on a general basis and on a project-specific basis.

6. In designing and implementing evaluation studies, maximum use should be made of host country skills and resources, such as local universities and consulting firms.

7. Achieving maximum utilization of evaluation findings in similar projects and programs in other countries requires an information system capable of collecting, storing, matching, retrieving, and disseminating experiential data.

THE PRECONDITIONS FOR EVALUATION

The key element both in project design and evaluation is the establishment of a logical framework for the project. The logical framework assists the designer to structure the project design in the following manner.

1. *To define a causal hierarchy of project inputs, outputs, purpose, and higher goal in measurable or objectively verifiable terms.*

Project Outputs are defined as the specifically intended kind of results (as opposed to their magnitude) that can be expected from good management of the inputs. Outputs are also defined as the preconditions necessary to the achievement of the project purpose.

Example: Manpower, training, machinery and building materials (inputs) can be managed to produce an irrigation network, trained operational staff, a water utilization schedule, and a user rate scale (outputs).

Project Purpose is defined as the primary reason for the project, i.e., the result which is expected to be achieved or the problem which is to be solved if the project is completed successfully and on time.

Example: An irrigation network and associated facilities and services (outputs) are intended to produce increased per hectare yield (project purpose).

Sector/Program Goal is defined as the programming level beyond the project purpose, i.e., the next higher objective to which the project is intended to contribute.

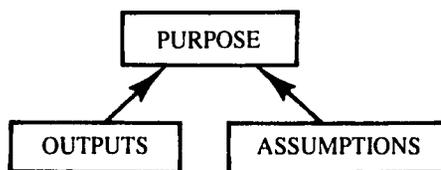
Example: Increased per hectare yield (project purpose) is intended to result in expanded exports of agricultural crops (sector goal).

2. *To hypothesize the causal (means-end) linkages between inputs, outputs, purpose and goal.* Each of these target levels must exist before the next can be achieved. The existence of one of the levels does not, however, guarantee that the next level will be achieved. Factors outside the project design (farmer attitudes, access to credit) may prevent achievement of the next level. Thus the causal relationship between any two levels in the hierarchy must be stated as a hypothesis, the external factors must be explicitly identified, and assumptions made about their behavior and probability of happening. Evaluation must then verify whether or not the hypothesis was realized.

3. *To articulate the assumptions about external influences and factors which will affect the causal linkages.* Assumptions are defined as external situations or conditions which must be assumed to exist or to be brought about if the project is to succeed, but over which the project management team has little or no control.

Example: Increased crop yield (project purpose) will contribute to expanded export of agricultural crops (sector goal) only if price and market conditions are favorable (assumption).

The concept of causality rests on the basic premise that each level in the causal hierarchy can be shown to be not only necessary but also sufficient to cause the next higher level to be achieved. Each causal linkage must be tested to assure that a given target level (e.g., outputs), in concert with the assumptions at that level, are necessary *and* sufficient to achieve the next level (purpose).



At the purpose-to-goal linkage, the causal model is basically the same although somewhat more complex in scope. This is because the necessary *and* sufficient



Kinds of Evaluation

There are several forms and types of evaluations used in development organizations. *Formative evaluation* is used when the project purpose is not readily definable in precise and explicit terms and when the strategy for achieving that objective is even less clearly understood. In such circumstances formative evaluation is used periodically to explore experience to date in order to permit managers progressively to sharpen the definition of the purpose and to formulate a viable strategy for achieving it.

Summative evaluation is used when the objective is clearly defined and there exists a high level of confidence in the strategy for achieving that objective. In this instance summative evaluation merely attempts to measure progress towards the objective.

The *goal attainment evaluation model* is used when the project or program has a single predominant objective. This model is widely known and used, it is relatively low-cost and imposes modest skill requirements.

The *systems evaluation model* is utilized for institution building and for complex projects and programs with multiple (and possibly competing or conflicting) objectives. It is not yet widely used because it is not well understood, and its cost is high in terms of the skills required.

Many institutions engage in an either/or debate about the comparative advantages of ongoing evaluation vs. ex post evaluation of completed programs and projects. However, both are needed. Evaluation of ongoing projects is a program management function intended to assure the efficient use of scarce resources, measure effectiveness in achieving planned project objectives, and confirm the continuing relevance of the project in the light of changing circumstances. Ex post evaluation of completed projects is a program policy and management function and is intended to give a final accounting to the sponsors and supporters of the project, identify possible follow-up opportunities for the host country, and derive lessons from past experience which can be applied to the design of similar projects and to the formulation of program criteria and policy.

conditions must include all projects as well as non-project initiatives (e.g., institutional, policy and other reforms) contributing to the goal.

4. *To establish the progress indicators which will permit subsequent measurement or verification of achievement of the defined outputs, purpose, and goal.*

Progress indicators are defined as pre-established criteria or measures of an explicit and specific nature designed to provide objective assessment of project progress. Progress indicators should be objectively stated so that both a proponent of a project and an informed skeptic would agree that progress has or has not been as planned. Pre-establishing objectively verifiable progress indicators and targets helps focus discussion on evidence rather than opinions.

The logical framework is primarily a project planning device. It also is used for re-examination of the original design of ongoing projects as a necessary prelude to evaluation. It sets the stage for determining whether or not the project outputs are being produced, whether these outputs in fact are serving to achieve the project purpose, and finally whether this achievement is making a significant contribution, as planned, to the higher goal.

The logical framework also is used to set the practical limits of responsibility of project management. Articulating the project planning assumptions in explicit and operational terms permits a clearer separation between manageable interests and those factors which appear to be beyond the control of the project management team. The input-to-output level is largely susceptible to managerial control with relatively few uncontrollable external factors. At the output-to-purpose level, the possibility of managerial control decreases while external factors become more important. At the purpose-to-goal level, the ability of project management to predict and control events usually is further diminished.

In addition to the logical framework elements noted above, the project design process must also contain such evaluation information and actions as:

- a. collection of baseline data;
- b. review of prior experience with similar projects elsewhere;
- c. provision for experimental, quasi-experimental, or other evaluation approaches.
- d. establishment of schedules for recurring evaluations with timing keyed to decision making.

For ongoing projects, evaluation is difficult if these elements of evaluation are not built into the original design.

Experience in evaluating ongoing development programs and projects has shown that for most projects it is not possible to evaluate progress toward established targets in a meaningful way until (1) the evaluator has considered any changes in the host country socioeconomic setting which may have significantly affected the project, and (2) the existing project design has been re-examined and clarified.

The first of these considerations, the assessment of changes in the host country socioeconomic setting, may well be demanding and difficult, requiring data collection and analysis as well as extensive discussion with all collaborating agencies. It includes such questions as:

- a. Changes in the nature and magnitude of the problem to which the project is addressed.
- b. The continuing validity of original feasibility data and estimates.
- c. Changes in physical and environmental conditions.
- d. Changes in demand and other economic variables.
- e. Changes in attitudes and other social variables.
- f. Changes in host country development policies and priorities.

When this assessment of changes in the project setting results in substantial redesign, it may not be possible to evaluate progress toward the new targets immediately. Depending upon the extent of the redesign and the revised implementation plan, it may be appropriate to postpone the evaluation for a year or more.

MEASUREMENT AND VERIFICATION

If the project design provides for the evaluative preconditions described above, then measurement and verification can be relatively simple and routine. Change can be measured by using the previously formulated progress indicators and drawing on the progress data collected during implementation. Program and policy evaluation, involving aggregations and inter-country comparisons of projects and programs, generally follows the same basic evaluative process as is followed in project evaluation.

Retrospective verification includes careful examination of the nature of the processes through which the hypothesized causal linkages occurred or failed to occur. Verification and validation of the hypotheses might be approached through direct observation, the use of proxy or surrogate measures, examination of differences in impact among target subgroups within the treated areas and of the independent variables associated with those differences, and statistical techniques such as multivariate analysis.

Verification can also be approached by formulating alternative hypotheses to explain changes. If none can be developed or if plausible alternative explanations can be disproven, this suggests at least partial validation for the original hypotheses. Conversely, if the alternative hypotheses can be supported to the point that they become persuasive, then the original hypotheses lose credibility. While such retrospective methods would appear less satisfactory than the classical experimental design, the operational realities of a development assistance program usually favor the former.

Evaluation findings should be reviewed to establish implications and conclusions for further action. The review should include all interested parties: the host country, the implementing agents, and all donors. The climate should be collaborative and constructive, not threatening. Replanning, not recrimination, is the intent.

ROLES AND RELATIONSHIPS

Two closely related pairs of issues require thoughtful consideration.

1. Should the evaluation responsibility be centralized (i.e., performed by a central evaluation unit) or decentralized? Since the critical criterion is that responsibility for conducting evaluations should be located functionally and organizationally as close to the decision-making point as possible, application of this criterion suggests that:

- a. Evaluation of progress and of continuing relevance of individual projects be decentralized to the project management team.
- b. Evaluation of experience to formulate policy and program criteria be done by a centralized unit concerned with policy and program coordination.
- c. Evaluation of programs and groups of projects for program management purposes be done at an intermediate level, e.g., regional or technical offices.

2. Should evaluations be conducted by in-house staff (by those with program or project implementation responsibilities) or by outside experts? Operational experience shows that project managers can evaluate progress and the continuing relevancy of their own projects candidly and objectively if there is a climate of constructive inquiry rather than of recrimination. There are simple criteria to guide the choice.

IN-HOUSE

- a. Greater knowledge of the environment and of specific operations is required.
- b. Direct and immediate feedback of findings into replanning.
- c. Lower unit cost—broader evaluation coverage.

OUTSIDE EXPERTS

- a. Disinterested objectivity is paramount and must be demonstrated.
- b. Greater and more recent technical knowledge is required.
- c. Evaluator must be free from operational workload for extended period.

Basically, it should not be an either/or choice. Both kinds of evaluation are needed, and a combination of in-house and outside experts often provides the best arrangement.

Irrespective of the extent to which program evaluation is decentralized, there is a need for a central program evaluation unit to assure that evaluation is occurring, is professional and rigorous, and addresses operational concerns. It would also make certain that evaluation components are built into programs and projects, that effective methods and techniques are available, understood, and used skillfully, and that findings are channelled into replanning and decision making.

The services which such a central program evaluation unit might give to evaluators in the organization would include: the coordination of evaluation activities and the preparation of annual evaluation plans; advice on, and participation in evaluations; administration of evaluation contracts and consultant services; improvement of methodology, guidance, instructional material; training in evaluation methodology; and a central evaluation document and reference service.

Appendix

The Logical Framework

The Logical Framework

1. A key element in project planning and evaluation is the establishment of a logical framework for the project design which:

a. Defines project inputs, outputs, purpose, and higher goal in measurable or objectively verifiable terms.

b. Hypothesizes the causal (means-end) linkage between inputs, outputs, purpose, and goal.

c. Articulates the assumptions (external influences and factors) which will affect the causal linkages.

d. Defines the indicators which will permit subsequent measurement or verification of achievement of the defined outputs, purpose, and goal.

A diagrammatic outline of the logical framework appears as Figure 2 below.

2. The logical framework methodology embodies the concept of causality; i.e., the causal linkage or hierarchy in which resource inputs are intended to produce outputs, outputs are expected to result in the achievement of project purpose, and project purpose is expected to contribute substantially to the higher goal. The concept of causality, in turn, rests on the basic premise that each level in the hierarchy can be shown to be not only necessary but also sufficient to cause the next higher level to be achieved. Since each causal linkage is subject to external factors beyond the control of project management, each linkage must be tested to

assure that a given target level (e.g., outputs), in concert with the assumption at that level are necessary and sufficient to achieve the next level (purpose).

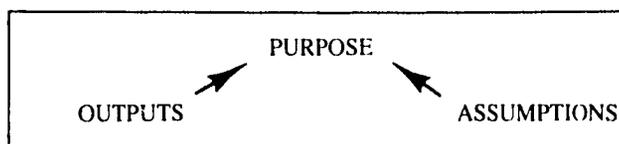


Figure 1

3. The logical framework is primarily a project planning device. It also is used for reexamination of the original design of ongoing projects as a necessary prelude to evaluation; i.e., it sets the stage for determining and validating whether or not the project outputs are being produced; whether these outputs in fact are serving to achieve the project purposes; and finally whether this achievement is making a significant contribution, as planned, to the higher order goal.

4. The logical framework also establishes the practical limits of responsibility of project management. Articulating the project planning assumptions in explicit and operational terms permits a clearer separation between manageable interests and those factors which appear to be beyond the control of the project management team. The input-to-output level is largely susceptible to managerial control with relatively few uncontrollable external factors. At the output-to-purpose level, the possibility of managerial control decreases while external factors become more important. At the purpose-to-goal level, the ability of project management to predict and control events usually is further

diminished. In evaluating project progress, it is necessary to examine the original planning assumptions about the role of external factors and to validate the hypothesized means-end linkages.

Logical Framework Characteristics and Limitations

1. All aspects of project planning (i.e., the formulation of targets, causal linkages, indicators, and assumptions) are defined by the project planners and are project-specific. Similarly, the degree of rigor and the level of effort in collecting and analyzing data for the evaluation are determined by the person/committee conducting the evaluation and are project-specific.

2. The logical framework methodology does not assure that the project is optimal; i.e., that the project directly addresses the most critical constraint to goal achievement, and is the most effective means for overcoming that critical constraint unless the planners and/or evaluators choose to explore alternative approaches.

3. The logical framework methodology gives no guidance on questions of equity or benefit incidence such as equitable income distribution, employment opportunities, access to resources, popular participation in decision-making and in the fruits of development projects unless such aspects have been explicitly included in the statements of goal or purpose. Guidance on benefit incidence policies and criteria has been issued periodically and will continue to be developed and disseminated.

4. The logical framework methodology is programmatically and technically neutral. It gives no guidance on proven strategies and techniques, cost and feasibility of replication, effects on the environment, concentration on key problem areas, reliance on the private sector, etc.

5. The logical framework methodology permits, but does not require, cost/benefit and cost/effectiveness analysis.

6. The logical framework matrix can be modified by the user for special circumstances; e.g., one or more horizontal rows can be added to provide for intermediate subsectoral goals. (See the *Logical Framework, Modifications Based on Experience, November 1973*.)

7. Further description and instruction concerning the logical framework methodology is found in the following, available from AID:

a. The Project Evaluation Guidelines, Supplement I, Third Edition, August 1974.

b. The Logical Framework—Modifications Based on Experience, November 1973.

c. AID Use of Development Indicators—A Progress Report, March 1974.

d. A glossary of terms associated with the logical framework methodology is found at the end of this section.

Content of Logical Framework

1. Goal-Narrative Summary

a. Goal is a general term characterizing the programming level beyond the project purpose; i.e., the next higher objective to which the project is intended to contribute. It provides the reason for dealing with the problem which the project is intended to solve. Goal denotes a desired result to which an entire program of development may be directed. Goals are established at top program management levels. Project managers need to understand these programming goals even though their contribution in formulating them may be limited.

b. Generally, a goal is not achieved by one project alone, but is established with the intent that success in a variety of projects and nonproject activities will be necessary for its achievement. In this respect, the relationship between goal (the end) and project purpose (the means) is causal and partial. Causal relationships become more direct and complete when descending to the *output* and *input* levels. The establishment of a goal is thus only one final stage in a logically progressing series of hypotheses:

(1) *If* this goal is desirable, *then* what project purpose will be necessary to achieve it?

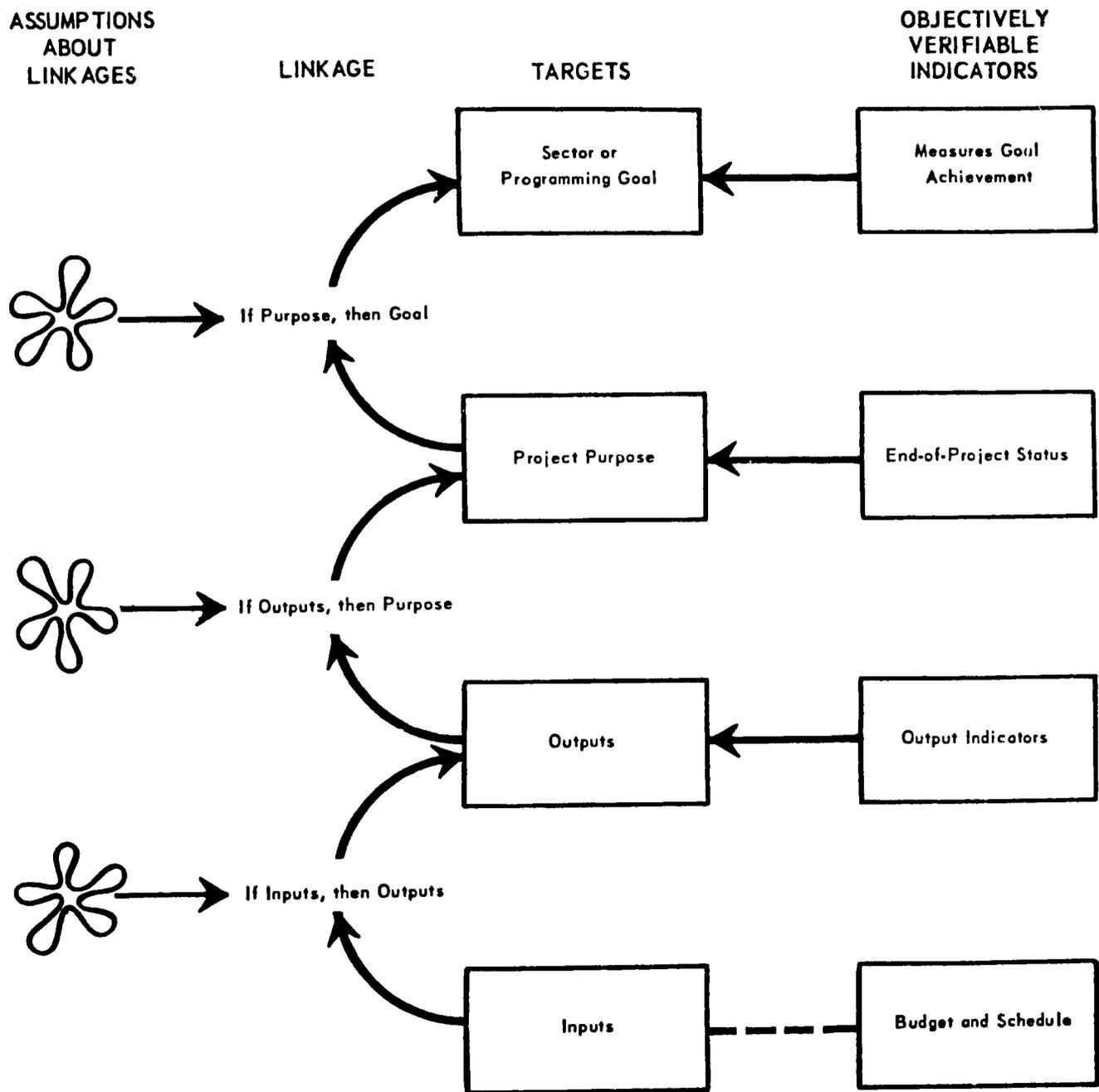
(2) *If* this project purpose will assist goal achievement, *then* what outputs will be necessary to achieve the project purpose?

(3) *If* these outputs are to be provided, *then* what inputs will be required?

2. Goal-Objectively Verifiable Indicators

The indicators of goal achievement may be quantitative, qualitative, or behavioral, or a mixture of these criteria. Satisfactory measures of achievement are those which indicate a realistic *causative* relationship between project purpose and goal and confirm that the project purpose contributes to the achievement of the

THE LOGICAL STRUCTURE OF A PROJECT



**PROJECT DESIGN SUMMARY
LOGICAL FRAMEWORK**

Life of Project:
From FY _____ to FY _____
Total U.S. Funding _____
Date Prepared: _____

Project Title & Number: _____

NARRATIVE SUMMARY	OBJECTIVELY VERIFIABLE INDICATORS	MEANS OF VERIFICATION	IMPORTANT ASSUMPTIONS
Program or Sector Goal: The broader objective to which this project contributes:	Measures of Goal Achievement:		Assumptions for achieving goal targets:
Project Purpose:	Conditions that will indicate purpose has been achieved: End of project status.		Assumptions for achieving purpose:
Outputs:	Magnitude of Outputs:		Assumptions for achieving outputs:
Inputs:	Implementation Target (Type and Quantity)		Assumptions for providing inputs:

goal. Measurement indicators such as the number of local citizens taking part in an election, increased per capita income over a prior period, increased value of exports, and the number of job vacancies at a particular level in government and the private sector, provide a realistic picture of a situation at any given time. The scope of a single project will not usually be comprehensive enough to be the total cause of achievement of the goal. Other projects and nonproject factors may also have a significant influence on goal achievement.

3. *Goal-Means of Verification*

State the kinds and sources of data needed to support the indicators which have been cited as measures of goal achievement.

4. *Goal-Important Assumptions*

Achievement of the goal (and indeed the project purpose and outputs as well) is based on the expectation that certain events or actions, outside the scope of the project will occur. These external factors need to be stated clearly as important assumptions regarding goal achievement and evaluated periodically to assure their continued validity. "Increasing agricultural productivity," for example, may be a realistic goal. However, achievement of that goal may be dependent on motivating the farm labor force; establishing marketing regulations, distribution centers, and national price structure; and acts of God, such as weather, etc., factors clearly outside the design of the project. The degree of confidence that is placed on the assumptions about these factors depends on familiarity with the cooperating country, knowledge of the sector of concentration, cooperating country performance, etc. A project design is only as sound as the strength of its weakest important assumption. As the project is implemented and the hypothesized causal linkages are tested, the confidence level in the causality between purpose and goal should increase. If this does not occur, the evaluation process should then focus attention on the explicit assumptions.

5. *Project Purpose-Narrative Summary*

The project purpose is the specific desired result of the project, not merely the sum total of outputs. A well conceived project has an explicit defined purpose that contributes causally to the goal in a logical and direct manner. In turn, the combined effect of project outputs contributes in a logical and direct manner to achievement of the project purpose. This purpose represents the solution to a specific development problem and may be derived by inverting the statement of the problem into a statement of the appropriate solution.

6. *Project Purpose-Objectively Verifiable Indicators*

a. The statement of the end-of-project status conditions (EOPS) is a description of the set of terminal conditions that will exist when the project purpose is successfully achieved. This description takes the form of objectively verifiable indicators, either quantitative, qualitative, or behavioral in character which reflect the end of the project status conditions. In projects which have an institutional purpose, the end-of-project status conditions would include the actual performance of the institution, rather than its readiness (the latter would be output indicators). Indicators of institutional performance would include self-sufficiency, effectiveness in producing goods and/or services, efficiency, creativity, and initiative.

b. In projects that emphasize immediate accomplishments, the end-of-project status conditions expected often are direct results of project goods and/or services. Did the birth rate fall? Did exports rise? Did enough private enterprises (or cooperatives) survive to form a critical mass that will continue to grow without AID support? Do fewer children drop out of school as a result of the new instructional methods and textbooks? Did per hectare crop yield increase?

c. Targeted examples of conditions expected at the end of the project which are objectively verifiable include:

<i>Project Purpose</i>	<i>Conditions Expected at the End of the Project</i>
Upgrade marginally qualified and unqualified indigenous teachers already teaching in primary schools, and produce qualified new teachers for the expanding school system	1-100% of school have majority indigenous qualified staff by the end of 1975. 2-Unqualified teachers reduced (from 70% of 8,000 now) to 30% of 16,000. 3-25% of children that began grade one complete 7 years of school with a 75% pass exam record by 1975.

7. *Purpose-Means of Verification*

State here the sources and the specific types of evidence which will be used to verify conditions marking End-of-Project Status.

8. *Purpose-Important Assumptions*

As noted in C4, an assumption describes a situation or a condition which must be assumed to exist, if and

72 / The Logical Framework

when a project is to succeed, but over which the project management team may have little or no control. An example is: Increased crop yield (project purpose) will contribute to expanded export of agricultural crops (sector goal) only if price and market conditions are favorable (assumption).

9. Outputs–Narrative Summary

Project outputs are the planned results produced by the management of specific inputs. In analyzing project outputs, be aware of the distinction between the *kind* and the *magnitude* of the specific results that competent project management is reasonably able to produce. Producing trained cooperating country staff for certain key posts is an output. However, placing trained staff in a specified number of key posts within a particular time frame is an output indicator.

10. Outputs–Objectively Verifiable Indicators

The magnitude of outputs, targeted and expressed in a manner allowing verification, reflects evidence of successful completion of the managerial actions (input-output linkage) that were necessary to produce the output in the first instance. In the case just given, participant training would be the link. Examples of outputs and appropriate targeted output indicators include:

<i>Outputs</i>	<i>Output Indicators</i>
a. Trained indigenous personnel for key posts in Radio Correspondence (R/C) course;	a. Cooperating country personnel trained for, and assigned to 15 previously identified key posts by 1974;
b. Courses prepared and taped;	b. 18 courses prepared and taped by end of 1974;
c. Courses broadcast;	c. 18 courses broadcast by end of 1975;
d. Student enrollment;	d. 12,000-15,000 students enrolled by 1975;
e. Lesson marking system in operation;	e. Max. 4 weeks to return graded lesson to enrollees;
f. Teachers trained to pass an exam to qualify teachers at the 2nd level in a 3-level system;	f. 8,000 pass second level exam by end of 1975;
g. Research on effectiveness of R/C-trained teachers vs. untrained teachers.	g. Research report findings indicate R/C-trained teachers of higher quality than traditionally trained staff.

11. Outputs–Means of Verification

State the data source and kind of data for verifying each output indicator.

12. Outputs–Important Assumptions

- For general description of assumptions, see C4.
- Given outputs such as trained manpower (either through participant training or on-the-job training), a critical assumption may be that the government will formally establish appropriate positions and will budget funds to payroll them.

13. Inputs–Narrative Summary

Inputs are the goods and services provided by the Mission, the Bureau, the Office, other donors, and/or the cooperating country with the *expectation of producing certain definable outputs*. The inputs to a project may consist of personnel, equipment, commodities, training, funding, contract services, etc., in almost any combination. These inputs may be provided by the United States (directly or through contractors, participating agencies, or voluntary agencies), the cooperating country, or other donors. With respect to personnel the important factor is the services which each person is to perform rather than simply the assignment of an individual to the project; i.e., the fact that an advisor is at post is not a statement of the input expected from that advisor.

14. Inputs–Objectively Verifiable Indicators

For each element of the above input, list budget categories such as commodities (perhaps broken out into subgroups), participant training, advisory services (direct-hire or contract), and their quantities and approximate expenditure level.

15. Inputs–Means of Verification

This cell of the matrix may not have to be completed if inputs consist of AID Mission-furnished items for which AID records provide accounting. However, other inputs such as those by the cooperating country, voluntary agencies, and third countries, should have confirming data sources shown.

16. Inputs–Important Assumptions

Assumptions at the input level are usually limited to questions of whether the inputs will be available on time. Project designers may use this cell of the matrix to record “Beginning of Project Status conditions”; the project specific baseline conditions which are the obverse of the terminal or “End of Project Status condition.”

GLOSSARY

Assumption. An event or action which must take place, or a condition which must exist, if a project is to succeed, but over which the project team has little or no control. The explicit statement of such assumptions is an aid in reducing the uncertainty of the project's environment, and, by codifying the significant external factors, allows the project to be reevaluated and revised to allow for changing outside influences.

There are normally different assumptions for each level of the project design. For example, if the project purpose is to increase agricultural productivity through the development of a school of agriculture and the goal is to increase farm income to support local political stability, it probably would have to be assumed (a) at the goal level, that improved economic conditions will result in political stability, (b) at the purpose level, that the cooperating government will provide adequate budgetary support to the school after the completion of the project, and (c) at the output level, that there will be a sufficient number of students applying for places in the school.

BOPS. The Beginning-of-Project Status. (Use box D-4 of the logical framework matrix.) The baseline from which change will be assessed.

Conditions Expected At End of Project. See *End-of-Project Status (EOPS)*.

Development Hypotheses. "If outputs, then purpose" is the *project* development hypothesis. The hypothesis that project purpose will contribute to program or sector goal is the *program* development hypothesis. These are hypotheses because we are not certain of the causative relationship between the *if* statement and *then* statement. Projects should be supported only when informed judgment, based on the best available evidence, provides reasonable confidence that the *then* statement will be achieved.

End-of-Project Status (EOPS). The condition or situation which will exist if the project achieves its purpose; an objectively verifiable description of those conditions, in the form of measures, indicators, or proxies that will indicate the point at which the project purpose will be considered to have been achieved.

If we accept the premise that there is an "if-then" hypothesis relating outputs to purpose, it follows that we cannot measure outputs to find out whether or not we have achieved the purpose. The means of verifying achievement of project purpose therefore needs to be independent of, and different from, the means of

measuring outputs. Usually this will require the measurement of factors not under AID's control.

Goal. The term characterizing a programming level beyond the project purpose. It provides the *reason* for the project and articulates a desired end toward which the project efforts of AID (and the cooperating government) are directed. The rationale by which a project is undertaken should *ultimately* allow the project purpose to be linked to a goal (often at sector or program level) that is set out as part of the country strategy. However, it may at times be necessary to require setting intermediate goals that are both above the project level and below the level of impact discussed in the Development Assistance Plan (DAP). The goal normally deals with broad economic, social, and/or political problems. It may be measurable in quantitative terms, or it may be identified by qualitative and behavioral criteria.

GPOI. An acronym for Goal, Purpose, Outputs, Inputs.

Hypothesis. Webster's Third New International Dictionary defines hypothesis as "a proposition tentatively assumed in order to draw out its logical or empirical consequences and so test its accord with facts that are known or may be determined." To put it somewhat more succinctly, it is a statement in the form "if A, then B" where there is uncertainty about the causative relationship between the existence of A and the achievement of B. (See also *Linked Hypotheses*.)

Indicator. An explicit and objectively verifiable measure of results expected. Good project design must include preestablishing what will be measured or observed to demonstrate progress. Progress should be objectively verifiable so that both a proponent of a project and an informed skeptic would agree that progress has or has not been as planned. Preestablishing objectively verifiable indicators helps focus discussion on evidence rather than on opinions.

Indicators may be quantitative or qualitative. A quantitative indicator may be expressed as a single measure; e.g., 50 graduates during the 1972-'73 academic year; as a cumulative figure; e.g., 175 graduates since June 1968; or as a degree of change, usually a percentage figure or a ratio; e.g., 25% increase in the number of graduates per year between the 1971-'72 and 1972-'73 academic year.

In some cases, where quantitative measures are not possible, objective observation of a qualitative change may still provide a measure; e.g., working relations among cooperating-country personnel in extension service are significantly improved over 1 year, or,

students are participating more in unstructured classroom discussions and focusing less on rote memorization and regurgitation.

Sometimes it is not possible to measure a change directly as it is in the case of number of graduates per year, or yield per acre. In such cases, indirect or proxy indicators must be found; e.g., number of 6th grade graduates in a region as measure of literacy, or increased use of vaccine as a measure of improvement in the quality of livestock. When indirect measures are necessary, it is important to be sure the causal relationships that underlie them are verified. For instance, that a 6th grade certificate is an indicator of literacy in country x, or, that the particular vaccine is a sufficient condition to improve the health of livestock in region y.

Inputs. Inputs are the actions taken or goods and services (personnel, commodities, participant training, etc.) provided by the Mission, AID/W, other donors, and/or the cooperating country with the expectation of producing certain definable outputs. Thus, for example, with respect to personnel the important factor is the *function* which the person is expected to perform rather than simply the assignment of an individual. Inputs can usually be identified by asking, "What must be provided to produce the desired outputs?" It is an error, however, to use input language in a target statement; e.g., "To assist the Host Country to . . ." This tends to confuse cause and effect. In this case assistance would be the cause, and its requirements are not necessarily finite. Its effect, the target, should be explicit and have some definite end-status.

Linked Hypotheses. Using GPOI, the hypothesis is that achieving the expected results at each level of the GPOI hierarchy of means-ends relationships will lead to the planned results at the next higher level; that is, *if outputs are produced, then purpose will be achieved, and if purpose is achieved, then goal will be achieved, provided certain assumptions (external conditions and influences) operate as anticipated.*

Logical Framework. A summary, in matrix form, of project design, showing the results expected for each level of intent when a project is successfully completed. Results are expressed as objectively verifiable targets together with means of verification and controlling assumptions.

Matrix (Logical Framework). A summary worksheet for the analysis of a project design divided into four horizontal rows (for goal, purpose, outputs, and inputs) and four columns (for narrative, objectively verifiable indicators, means of verification, and important assumptions). Modifications may be made to suit local circumstances.

Measures of Goal Achievement (Indicators at Goal Level). The means of verifying the achievement (in either quantitative or qualitative terms) of the goal by means of appropriate indicators. Ideally, these might consist of increased per capita income over a given period, increased value of exports, percentage decrease of insurgent activity in a given area, etc.

Outputs. The specifically intended *kind* of results (as opposed to their magnitude) that can be expected from good management of the inputs provided.

Purpose. The primary reason for the project; i.e., the development which will be achieved or the problem which will be solved if the project is completed successfully and on time. The purpose expresses in quantitative or qualitative terms that developmental change which is to be created or accomplished with a view towards influencing the solution of a country or sector problem.

Target. An explicit and objectively verifiable statement of results expected within a specific time period; e.g., 100 tons/year in 1975, enabling legislation passed by 1972, 17 reports requested and completed by 1973.

We use the term *target* to specify the desired and product and any level of intent; i.e., output, purpose, goal. *Target* means performance standard. Target should contain at least three dimensions where feasible: magnitude, target area or audience, time.