FOR AGENDA

EBAP/92/166

CONTAINS CONFIDENTIAL
INFORMATION

December 17, 1992

To: Members of the Executive Board

From: The Secretary

Subject: Establishing an Evaluation Office in the Fund

This paper provides background information to the statement by the Managing Director on the establishment of an evaluation office in the Fund which was circulated as BUFF/92/141 (12/8/92), and is tentatively scheduled for discussion on Friday, January 22, 1993.

Mr. Prust (37884) is available to answer technical or factual questions relating to this paper prior to the Board discussion.

Att: (1)

Other Distribution:
Department Heads

CONTAINS CONFIDENTIAL INFORMATION

INTERNATIONAL MONETARY FUND

Establishing an Evaluation Office in the Fund

Report by a Task Force 1/

December 16, 1992

Contents

		Page
I.	Introduction	1
II.	Rationale for an Evaluation Office	1
	1. What is evaluation?	1
	2. Principal benefits of an evaluation office	2
	3. Possible reservations	4
III.	Current Appraisal Activities in the Fund	5
	1. Routine assessment of performance under Fund-supported programs	6
	2. Studies called for under the conditionality guidelines	7
	3. Annual SAF/ESAF papers	11
	4. Topical policy papers	12
	5. Technical assistance	13
IV.	Evaluation Experience of Other Institutions	14
	1. The World Bank	14
	2. Other institutions	17
V.	Characteristics of an Evaluation Office in the Fund	19
	1. Organizational status	19
	2. Scope of work	21
	3. Evaluating Fund-supported programs	22
	4. Evaluating technical assistance	25
	5. Preparation, discussion, and dissemination of findings	26
	6. Absorption	28
	7. Staffing requirements	28
VI.	Conclusions and Recommendations	30

1/ Chaired by P.R. Narvekar, the Task Force included Bahram Nowzad (from February 1992) and Jacques J. Polak (from May 1992), both as consultants working part time, and Jim Prust (from April 1992). The report was completed at the end of June 1992 but some changes, mainly editorial or designed to improve clarity, have been made since then.

Contents

		Page
Appendix I.	The World Bank's Evaluation Procedures and Experience	35
Appendix II.	Evaluation Practices of Other Institutions	46

I. Introduction

Calls for the establishment of an evaluation office in the Fund go back at least to the early 1970s when such offices were envisaged or created in many other multilateral financial institutions, and in recent years such calls have increased. Early this year, management set up a Task Force to study the matter.

In preparing this report, members of the Task Force have had informal discussions with individual Executive Directors, some World Bank Executive Directors, senior Fund and Bank staff, and the directors and other officials of the evaluation units of the World Bank, the African Development Bank, the Asian Development Bank, the European Bank for Reconstruction and Development, the Inter-American Development Bank, the OECD, the U.S. Agency for International Development, the General Accounting Office of the U.S. Congress, the U.K. Overseas Development Administration, the Swedish International Development Authority, and the Japanese Overseas Economic Cooperation Fund. Evaluation units in some other organizations were contacted through correspondence.

This report begins, in Section II, with a general statement of the main considerations regarding an Evaluation Office in the Fund. Section III provides a summary of existing evaluation-type work in the Fund. Section IV reports on the evaluation practices and procedures of the World Bank and some other organizations; these are described in greater detail in Appendices I and II, respectively. Section V sets forth in broad outline what could be the main characteristics of an Evaluation Office in the Fund. The Task Force's conclusions and recommendations are presented in Section VI.

II. Rationale for an Evaluation Office

Practically all multilateral and official bilateral lending institutions include a separate evaluation office. The Fund, on the other hand, has never had a distinct entity responsible solely for evaluating its activities.

This absence is not in itself sufficient justification for the creation of an evaluation office in the Fund; in several respects the Fund differs significantly from other multilateral financial institutions. Rather, the case for an office must turn on whether its establishment would enable the Fund to better carry out its mandate and responsibilities, and thus to better serve the needs of its members.

1. What is evaluation?

Evaluation can be viewed as a critical *ex post* examination of the activities of an institution that is distinct from--though a complement to--

the ongoing monitoring that takes place during program or project implementation. To be credible, evaluation must be carried out by qualified persons who are independent of the formulation and implementation of the activities being assessed and whose impartiality cannot be questioned; and through channels that are not the same as those through which the work being evaluated was carried out. To be effective, procedures must exist for incorporating evaluation findings into future operations and policy formulation.

A considerable amount of work that has some of the characteristics of evaluation is already undertaken in the Fund. Area departments undertake ex post reviews of programs; PDR prepares periodic conditionality reviews; and other departments (in particular, Research and Fiscal Affairs) conduct ad hoc studies of specific aspects of Fund policies and practices (such as the Fund's recommendations on exchange rate policies in developing countries, or the experience with programs in Eastern Europe). But, in the opinion of many, the credibility of these efforts (which are reviewed in Section III) is put into question, in particular by the fact that to a large extent they consist of reviews by the staff of its own work. It should be stressed that there is no allegation of impropriety or of deliberate suppression of unfavorable findings. Rather, the claim is that review work which is prepared, commented on, and cleared through the same channels as the work which is being reviewed is not the best way to obtain independent, critical judgments. If the particular staff doing the review work also happen to have been involved with the work under review, the problem is compounded.

To meet this criticism, evaluation of the activities of the Fund would need to be carried out **independently** of the regular, mainstream work structures of the institution, and **outside** the normal lines of responsibility and reporting. This may best be done by entrusting these functions to an office whose main, or sole, responsibility would be to evaluate. Independence and focused responsibility would avoid actual or apparent conflicts of interest and, therefore, give credibility to the evaluation process.

An evaluation office would not have exclusive responsibility for evaluation in the Fund. Keeping work experience continuously under review is in any event a vital function of senior staff in all departments. Moreover, for an evaluation office to execute its responsibilities effectively, extensive contacts with the departments of the Fund would be necessary, both to provide much of the primary raw material for the work of the office and to discuss findings at various stages in their preparation. Furthermore, many of the broader policy-oriented and cross-country studies that are now prepared in PDR, Research, and other departments would still be needed and would complement the work of an evaluation office.

2. Principal benefits of an evaluation office

For convenience, the benefits from establishing an evaluation office in the Fund may be classified under four, mutually reinforcing and interrelated, headings. Of course, some of the benefits mentioned may be obtained in various ways, of which establishment of an evaluation office is only one.

(a) Accountability. In a private corporation, profits, sales, and other similar measures provide its shareholders with an index of the success or failure of the company. In the case of an organization like the Fund, however, the picture is more complex as no market test is available. Profitability cannot be used as a guide. The "products" and services provided are more intangible, their impact more diffuse, and the accounting time-scale of an entirely different order of magnitude. In these circumstances, the success of Fund programs must be, to a substantial extent, a matter of judgment. The essential role of an evaluation office would be to provide frank and independent judgments on the performance of the institution to the management and the Executive Board. Moreover, by providing such assessments, the office could improve the credibility of the Fund with the media and the public at large, and solidify the trust and support of member governments.

(b) Learning from past experience. The Fund has a rich and varied reservoir of experience which, judiciously tapped, can yield relevant lessons for future operations and approaches.

First, Fund-supported programs have in many instances not been successful, or at least they have not fully achieved their intended results: frequent slippages in programs, continued use of Fund resources over long periods of time involving a succession of programs, and the accumulation of arrears in some countries, are obvious examples. An independent evaluation of past experience could shed fresh light on the reasons for failure or for results that fall short of objectives. By the same token, evaluation could help identify and explain the examples where Fund recommendations and Fund-supported programs were substantially successful. In brief, evaluation could explain, based on close review of individual cases, **what works, what does not work, in what circumstances, and why**.

Second, assessments of past experience in all its major aspects (goals, performance, implementation, etc.) could provide pertinent information that would assist the staff, management, and the Executive Board in their decision making. (Such information would also bring about greater "parity" between what information is available to the Fund and Bank boards, respectively. Where a Bank structural adjustment or sector loan is closely associated with a Fund arrangement, the Bank's evaluation reports often contain some discussion of the aims and successes of the Fund arrangement. The information and appraisal thus provided may go beyond anything the Fund Board receives from the staff; an evaluation office in the Fund could give the Board its own source of critical appraisal of Fund programs, as well as of any associated Bank program.)

Third, a centralized and systematic assessment of policies and recommendations could promote **consistency** (with due regard for differing circumstances) in the application of conditionality over time and across countries.

Fourth, evaluation could become a vehicle for the Fund to study and learn more about the **longer-term impact** of its policy recommendations on member countries. The Fund frequently portrays the programs it supports as transitions to a more efficient economy which will lay the groundwork for higher growth and a viable balance of payments. Evaluation of the longer-term impact of Fund-assisted programs could help assess whether the adjustments encompassed in these programs lead to sustainable improvements in the economies in question.

(c) "Healthier" institution. The self-critical look that independent evaluation would bring could guard against inward-looking tendencies in analysis and policy recommendations and thus help foster healthier professional attitudes in the institution. The major activities of national civil services are routinely subject to political debate and to public discussion by a variety of interested parties. The Fund has become a more "open" institution over the years and its activities attract a great deal of outside comment. Nevertheless, the need remains to guard against any tendency for attitudes and practices to become inbred.

Independent evaluation could also encourage constructive discussion among the staff and, in particular, could provide an opportunity to air views which did not prevail at the time that decisions were made on the matters under review.

(d) Institutional memory. Independent evaluation could contribute to developing the institutional memory of the Fund by broadening its information base. While, compared to many other institutions, the Fund maintains excellent documentary records, it is less easy to gauge the track record of the Fund's operational activities, especially as regards specific countries or activities. An evaluation office could, for instance, provide on-line computerized summaries of its assessments and thus enable easy access to the record of past experience. Its reports and studies would become part of the institution's basic documentation.

3. Possible reservations

As stressed elsewhere in this report, to be effective an evaluation office would need to have the support of the Board and management, adequate staffing, and a mechanism for ensuring the proper absorption of its findings. Even if these conditions were fulfilled, it cannot be taken for granted that the potential contribution of an evaluation office to the Fund's effectiveness would be realized.

First, the experience with evaluation units in other institutions is only partially relevant to the Fund. Fund programs encompass the entire economy of a member country and a complex interplay of economic and political considerations. Their evaluation inevitably raises difficult judgmental issues in ways that evaluation of projects or even of sector programs may not. Whatever difficulties there are in analyzing and drawing lessons from the record of the Fund's experience would not disappear with the establishment of an evaluation office. Where economic relationships are imperfectly understood and changeable; where feasible economic policies must pass the test of political acceptability; where experience is diverse; where the Fund, like any other creditor, is inevitably taking some risk in making its resources available, the drawing of clear and practically applicable lessons will not be easy for anybody.

Second, the prospect of ex post analyses of their best efforts, conducted with the benefit of hindsight, and without a full appreciation of the imperfect knowledge, uncertainties, and difficult choices that had to be faced in negotiating programs or formulating policy advice, could arouse concern among the regular staff. Unless these concerns were allayed, there would be a risk that the needed cooperation, and receptiveness to recommendations of the evaluation office, could be jeopardized. Another danger is that staff members could become overly cautious in their work so as not to appear in a negative light at a later date, even though, as it must be emphasized, the purpose of evaluation is to learn from the past and not to assign blame.

Third, timeliness is key to the effectiveness of evaluation and the experience of other institutions in this regard has been less than satisfactory. But even assuming a much shorter lag of evaluation in the Fund, the benefits of purely backward-looking evaluation could be limited in circumstances where the principal issues facing the Fund are changing rapidly (as happened, for instance, at the onset of the debt crisis or in the wake of changes in Eastern Europe and the former Soviet Union).

Finally, creating an evaluation office could--without adequate safeguards in determining its work program--encourage outside groups with specific and narrow agendas to pressure the Fund to conduct evaluations focusing on their particular interests, which may lie outside the core mandate of the Fund.

III. Current Appraisal Activities in the Fund

Even without an evaluation office, the Fund engages in a variety of activities that look back at experience and attempt to draw useful inferences from it. The review of the principal recent examples of such activities presented in this section attempts to highlight some major aspects that are relevant to the evaluation function: On what subject do they focus? How effective and candid are they in drawing lessons from past experience? How credible are they?

The various appraisal activities described below show considerable differences with respect to these questions. But two general conclusions stand out: (i) Much of this work sheds light on the contents of Fund programs, members' experience in their execution and the response of countries' economies, and thus helps to draw general conclusions about which policy prescriptions work; but it is not intended to evaluate the performance of the Fund as it deals with the problems of individual member countries. (ii) The appraisal activities described in subsection 2 were largely carried out in the same departments as had substantial responsibility for the work under review; 1/ efforts are, however, made to ensure that the same individuals do not review work in which they have been personally involved. There is no suggestion that internal review may not, in some instances, be just as critical as review by an Evaluation Office (see the case of Egypt in subsection 2b, below). In general, though, it appears that the criticism contained in these various review papers has been more directed at member country governments than at the work of the staff or of the Fund as a whole.

It may also be noted that the Fund has only limited provision for the dissemination of the findings of the various studies and their incorporation into future operations. Papers prepared for the Board are, of course, distributed widely but staff members often cannot take time to study these often lengthy papers, or to attend the Board meetings or seminars where they are discussed. Periodic decisions by the Board to let its 1979 guidelines on conditionality stand unchanged also mean that the staff is not alerted to the shifting weight put on various policy elements in these guidelines. While PDR has had the task of ensuring consistency of treatment across members, and there is an active pass-through of information by word of mouth especially within area departments, the Fund has not systematically brought the operational staff up to date on the findings concerning the Fund's experience to ensure that they are taken into account in formulating future operations. However, recent changes in the mandate of certain divisions within PDR have been made partly with these issues in mind.

1. Routine assessment of performance under Fund-supported programs

Virtually all Fund-supported programs are given a follow-up assessment in either the next staff report for the Article IV consultations or in the papers for subsequent use of Fund resources. These assessments--sometimes presented in a special section of the report, sometimes in the staff appraisal, and sometimes dispersed through the documentation--vary in their degree of thoroughness, clarity, and candor. But--given the purpose of the reports of which they form a part and perhaps the fact that they are prepared by the department largely responsible for the program under review--their focus is mainly on the country's performance, not on the major

---

1/ An interesting exception prepared some 30 years ago, and thus of largely historical interest, is an evaluation of the Fund's relations with Colombia, 1957 to 1961, written by a staff member of the Research Department: Anne W. Romanis, Stabilization and Development: The Colombian Case (January 7, 1963).

aspects of the program itself and the recommendations given by the staff. They are no substitute for the thorough and candid assessments that genuine evaluation would require and, indeed, are not intended to be. In addition to the above, various papers are prepared from time to time on an ad hoc basis in area departments reviewing the Fund's relations with particular countries.

2. Studies called for under the conditionality guidelines

Pursuant to Guideline 12 of the 1979 Board decision on conditionality, the staff have prepared nine studies in the last 12 years examining the Fund's experience with upper credit tranche conditionality. The aim of these studies as specified in the above Guideline was "to evaluate and compare the appropriateness of the programs, the effectiveness of the policy instruments, the observance of the programs, and the results achieved." The purpose is thus broader than an appraisal of conditionality as such. The studies attempt to formulate broad generalizations about the content and effectiveness of programs supported by the Fund.

All but one of these studies were presented as cross-country reviews, each based on a study of all credit tranche arrangements of a particular generation; for example, the 1991 review derives its findings from all arrangements approved from January 1985 to June 1988; only the 1989 conditionality review presented a detailed study of a small sample of cases, although individual case studies did underlie some of the other reviews.

a. Cross-country reviews

The purpose of the 1991 review, as well as of earlier cross-country reviews, is stated to be to draw "general conclusions about the effectiveness of Fund-supported programs in meeting their objectives" and "to measure compliance with policy targets and to assess the extent to which program objectives were achieved with respect to variables such as growth, inflation and the balance of payments." 1/ To draw out this information, the general format of these reviews is to present tables on certain features of the economies studied (growth, inflation, current account balance before and after the program; program targets and realization), with averages for all countries and subgroups; and policy measures prescribed and the extent to which they were implemented.

It is, of course, desirable to record how far countries were able to meet their objectives, but it is important to be aware of the limitations of what is being assessed. In the first place, although the adequacy of policies adopted to meet the chosen objectives is extensively discussed, no attempt is made to appraise the reasonableness of the objectives themselves. Do they reflect a sufficient effort or are they, perhaps, too ambitious? Such questions are not posed. Second, the meaning of the targets (or

---

1/ EBS/91/101, pp. 5,7.

objectives--the terms are used interchangeably) should be understood. With respect to the balance of payments (or the current account) they are typically the starting point of the staff's programming exercise, and may be regarded as the objectives of the program. But as regards the growth rate and the inflation rate (the two other variables that the staff traditionally assesses), the basis of the target figures incorporated in programs is often less robust. They tend to be the staff's--or the government's--best estimates (sometimes erring on the side of optimism 1/) of what the outturns will be, on the basis of the policies on which agreement could be reached and the assumed values of major exogenous variables. Achievement of targets derived in this way thus indicates merely that the staff and the authorities did a good forecasting job, or--as noted in the 1991 exercise--that the errors due to failure in policy implementation were roughly balanced in their effects by external developments.

Despite these limitations, the staff does draw some interesting indications by cross-classification of policies and results. For example, in 17 of the 44 arrangements in the 1991 study, countries met both the fiscal and the credit targets; of these, 13 did and 4 did not meet their growth objective. A further interesting observation, this one not found in the text, which is cast entirely in terms of outturn versus target: these 17 annual programs showed an average growth rate of 4.4 percent, while the 27 other cases had a growth rate of 1.3 percent (calculated on the basis of data in Tables 5 and 6 of EBS/91/101).

Findings such as these support the case that Fund programs, far from immiserizing, have had a beneficial effect on the countries that adopted them and then stuck to them; as such they are very germane to assessment of the Fund's work. They provide an indication of how particular policies are working in countries with Fund-supported programs. Nevertheless, there are limitations to the results which cross tabulation of a limited sample of results can yield. If one wants to assess the effects of given policies or exogenous factors--whether for all countries or only all program countries (such as the effect of the real exchange rate on the volume of exports, studied in the 1991 review)--there is no reason to limit the data base to the particular subset of countries (in this case only 21) covered by the conditionality review in any given year. Significant answers to such questions are more likely to be found using larger data sets and more sophisticated techniques, as has been done in a number of studies of the

---

1/ See Manuel Guitian, "Fund Conditionality: Evolution of Principles and Practices," IMF Pamphlet Series, No. 38 (1981), p. 38: "Frequently, targets are formulated with a view to influencing the actual results. This is particularly the case with expectations; the demonstration effects of the announcement of policy targets are important elements of adjustment strategies in that they often influence the formation of expectations. This argues in favor of setting "ambitious" (though perhaps "unrealistic") targets even when it seems, a priori, clear that their complete attainment is unlikely."

effect of Fund programs on growth, inflation, etc., which have been prepared by the staff in recent years. 1/

b. The 1989 case-study approach

For the 1989 conditionality review, the staff organized the material around 9 case studies. These case studies came closer than anything the staff has done in recent years to appraising the Fund's work on individual countries.

The country papers that constitute the 1989 review (which average about 40 single-spaced pages each) contain a rather detailed description and appraisal of each country's adjustment policies, the degree of success achieved and the quality of the staff work (and the performance of the Fund as a whole) in each case. Of the nine cases covered, typically over about a decade, three (Bangladesh, Chile, and Ghana) were on the whole successful and three more (Mexico, Morocco, and the Philippines), while starting out rather poorly in the first half of the 1980s, had taken a more satisfactory turn in recent years. The experience with Yugoslavia was also described as "mixed". Two (Egypt and Zambia) were obvious failures.

At first blush, the country studies strike the reader as exhibiting unusual frankness. A second look reveals that the frankness is rather selective. There is no hesitation in pointing out policy weaknesses on the part of the government. By contrast, to find any criticism of the quality of the work by the staff (for example, on the question of program design) one frequently has to read between the lines. It is relevant to note that, ultimately, weak program design cannot be blamed on the exigencies of negotiating agreements that are acceptable to both sides. Neither can matters of program implementation be considered the sole responsibility of national governments. The guidelines on conditionality state (guideline 7): "The Managing Director will recommend that the Executive Board approve a member's request for the use of the Fund's general resources in the credit tranches when it is his judgment that the program is **consistent with the Fund's provisions and policies and that it will be carried out**" (emphasis added). It is, therefore, incumbent on management and staff to address both the design and prospects for implementation of programs, and not to recommend programs for approval when either is considered inadequate.

1/ For example, Morris Goldstein and Peter J. Montiel, "Evaluating Fund Stabilization Programs with Multicountry Data: Some Methodological Pitfalls," IMF Staff Papers, June 1986, pp. 304-44; Mohsin S. Khan and Malcolm D. Knight, "Stabilization Programs in Developing Countries: A Formal Framework," IMF Staff Papers, March 1981, pp. 1-53; Mohsin S. Khan and Malcolm D. Knight, Fund-Supported Adjustment Programs and Economic Growth, Occasional Paper 41 (1985); Mohsin S. Khan, "The Macroeconomic Effects of Fund-Supported Program," IMF Staff Papers, June 1990, pp. 195-231.

The contrast in the treatment of national governments and the staff, respectively, can be illustrated by reference to the case study of Yugoslavia. For example, on the government, the paper noted: "Inadequate advances in structural areas...reflected weak implementation because of the strong resistance to change...policies often fell short of original intentions..." "The required cooperative effort of the republics and provinces often could not be sustained." "As the external position improved, the authorities' resolve to follow through with structural reforms waned." "During the period of enhanced surveillance in 1986-87, there was a reversal of policies in a number of areas."

As regards the staff's work, there appear to be only questions. For example, "The overall experience...raises the question of whether the policy mix should and could have had a stronger structural content." The "focus of fiscal policy may have been too narrow". Credit limits "might have been specified" to include currency revaluation effects.

To some extent, this disparity of language is inevitable. Whether the government carried out agreed policies or not is a fairly straightforward question; whether policies could have been different, and if so in what way, is an inherently more complex and contentious issue. Nevertheless, an appraisal that confines itself to the posing of questions, while eschewing firm answers, is not a helpful way of drawing lessons from past experience, and does not suggest an even-handed treatment of the Fund and national authorities, respectively: after all, the latter may have faced compelling factors which prevented policy from being implemented as scheduled.

In other instances too the analysis was not pushed to a point to yield useful conclusions. For example, in the cases of Zambia (1984) and Morocco (1985), the staff concluded only that "the conditionality may not have been entirely appropriate, reflecting the need to maintain the confidence of the international community and to avoid the accumulation of external arrears, including to the Fund" (pp. 24-5 of the summary paper (EBS/89/17)).

Only the case study on Egypt (which covered a single program) is quite explicit in its criticism, and it deserves to be recalled here in some detail. After negotiations stretching over nearly four years, an 18-month stand-by arrangement for Egypt of SDR 250 million was approved by the Board on May 15, 1987 and a first drawing of SDR 116 million, equal to Egypt's first credit tranche, was made. The arrangement was unusual in a number of respects. For the Fund, the prior actions taken were sufficient to justify Egypt's drawing her first credit tranche, with any further drawing requiring substantial further action. In the eyes of the Paris Club, however, Fund approval of an arrangement that appeared to foresee drawings in the upper credit tranches sufficed for the rescheduling of US$5.5 billion of arrears and debt. A further unusual feature was that, the Managing Director, concerned about the question whether the program would restore balance of payments viability, required that creditor governments reassure the Fund regarding Egypt's financing requirements. The program soon became inoperative as Egypt missed various performance criteria for end of June 1987.

The case study is outspoken in drawing three lessons from this experience, implying in effect that staff and management were wrong to have proposed, and the Board to have agreed to, the Egyptian arrangement:

(i) "Where problems are serious and longstanding, there is need for strong initial action to start the program on a sound footing..."

(ii) "When a large amount of Fund resources is to be provided up front, there is reason to seek strong initial policy action" (the implied question of phasing first credit tranche access is not touched on in the report).

(iii) "The Fund should not substitute assurances from other creditors for the strength of the program itself..."

Because of the total length of these case studies, a brief (29 pages) staff summary became the basis of the Board discussion. That paper's focus was again on the working of the Fund's conditionality in general--which the staff found broadly satisfactory--not on the quality of the work of the Fund in individual cases. Thus, while the staff may have learned a great deal by studying nine country cases in depth, the Board discussion did not scrutinize the individual cases; and, there was no reference at all to the one forceful disapproval in the background papers, that relating to Egypt 1987. In the Board discussion, several Directors suggested that studies of this character should be made by "an independent unit." The Chairman's summing-up noted that Directors "welcomed the case-studies approach," but also that they favored greater use of cross-country comparisons, using all cases rather than just a selected sample. Accordingly, the next review (in 1991) returned to the cross-country approach.

3. Annual SAF/ESAF papers

Since 1988, the Fund has reviewed each year the program experience under SAF and ESAF. In many respects, the staff papers prepared for these exercises are comparable to those dealing with stand-by and EFF arrangements discussed in the previous section, but there are some important differences.

First, since major features of SAF and ESAF (access, conditionality, duration) remain under active review, the papers (and the Board discussion) have a more operational character. Perhaps because there is greater homogeneity among the SAF/ESAF countries (many of which are in Sub-Saharan Africa) than among the SB/EFF countries, the papers seem to be more successful in drawing lessons not only about the countries' performance, but also about the Fund's, and thus carry more of an evaluation flavor.

Second, the priority given to growth in SAF/ESAF programs provides an intuitively direct measure of success that is lacking in the reviews of stand-by and EFF programs where the primary emphasis is on payments adjustment: growth of 4 percent or more per year can generally be rated as reasonably good, but whether a current account deficit of a given percent of GDP is good or not depends on assumptions about current and future financing that may be much more disputable. With a relatively long period of observation (also associated with the SAF/ESAF medium-term approach), conclusions on the growth effect can be drawn without the need for excessive statistical concerns. Thus, a statement that "the median growth rate over the last 3-4 years under SAF/ESAF for the 22 countries under review has now reached 4.2 percent per annum, compared to a median of about 2.6 percent in the three-year period before SAF/ESAF support" (EBS/91/109, p. 8) appears both statistically acceptable and quite relevant to the experience with the facilities. With this finding established, much less is made of questionable outturn/target comparisons for inflation and external sector developments than in the conditionality papers.

For operational, and perhaps also for other, reasons periodic SAF/ESAF papers may need to be continued for as long as the facilities, or others like them, are operative. They are no substitute, however, for more detailed assessments of the Fund's experience with individual countries. As most ESAF arrangements may rather soon reach their three-year (or in some cases four-year) ends, full-fledged evaluation of the experience under these arrangements would be valuable.

4. Topical policy papers

From time to time, papers analyzing various topical aspects of Fund programs have been prepared for discussion (often in seminar format) in the Executive Board. Three outstanding such papers of recent vintage are listed below.

In 1986, the Board discussed an extensive paper on Theoretical Aspects of the Design of Fund-Supported Adjustment Programs (published in 1987 as IMF Occasional Paper No. 55). That paper explained the Fund's standard financial programming framework and the manner in which a variety of policy options could be incorporated into it. Exchange rate policies received considerable attention in that connection, and they were made the exclusive subject of the second paper which was on Exchange Rate Policy in Developing Countries: Some Analytical Issues (later published as Occasional Paper No. 78, 1991). This was discussed in a Board seminar in 1990. Late in the following year, Michael Bruno (the eminent economist and former Governor of the Bank of Israel) was commissioned by the European I Department to appraise the stabilization and reform programs of five Eastern European countries. The paper came out as WP/92/30: Stabilization and Reform in Eastern Europe: A Preliminary Evaluation.

Papers of this general nature--which the staff of the functional departments will no doubt be requested to produce from time to time in the future as well--and the discussion to which they give rise, can contribute to improved performance of the Fund in the design of programs and in its policy advice in consultations. If an evaluation office is established, it too might sometimes prepare papers of a general nature, especially after it had gained some experience in evaluating Fund operations.

5. Technical assistance

Technical assistance is an important and expanding activity of the Fund. In FY 1992, it absorbed about 200 man years, and accounted for about one quarter of the total resources devoted to country-specific work.

Technical assistance is typically provided either through staff missions to the host country or through assignments to the country of experts under contract to the Fund for varying periods. Increasingly, assignments of experts under technical assistance programs are externally financed, principally by the United Nations Development Program (UNDP) and the Japan Administrative Account. Such assignments are expected to account for approximately 20 percent of the total technical assistance work in FY 1993.

Review and assessment of technical assistance requests and activities is carried out in various ways. 1/ The Fund-wide Technical Assistance Committee, established in 1991, has instituted procedures to formalize discussions between area departments and technical-assistance-providing departments on the allocation of assistance both regionally and country-by-country. Experts in the field are generally required to remain in regular contact with Fund headquarters concerning their assignment and staff may from time to time visit the country to assess progress. Terms of reference and post-mission reports are reviewed by the relevant area and functional departments. In addition, a few of the projects jointly undertaken with the UNDP have been subject to tripartite evaluation--by the Fund staff, UNDP representatives, and the government of the host country.

These various procedures involve some element of review and appraisal regarding both the substance of the advice provided and the manner of its delivery. However, they cannot be regarded as--and are not intended to be--full-scale evaluation exercises.

1/ Until 1990, Annual Reports on Technical Assistance provided information on all the Fund's technical assistance activity for the year as well as summary accounts of individual projects.

IV. Evaluation Experience of Other Institutions 1/

1. The World Bank

The Bank's evaluation function is generally regarded as the most developed among the multilateral financial institutions. The evaluation system, which has evolved over more than two decades, is designed to promote credibility and accountability, to learn from the lessons of past experience, and, in these ways, to enhance the Bank's operational effectiveness and service to members.

The Operations Evaluation Department (OED) was established in 1973 as a unit independent of the Bank's mainstream structure; it is headed by the Director General, Operations Evaluation (DGO), who has the rank of Vice President. He is appointed by, and reports directly to, the Board and, when his term expires, he may not re-join the staff. In addition to heading the OED, his responsibilities are to oversee the adequacy of the Bank's overall evaluation function, to assess the effectiveness of operations, and to make recommendations designed to enhance the Bank's efficiency and responsiveness.

A Joint Audit Committee (JAC) of the Board oversees OED's activities, discusses OED reports, scrutinizes and recommends to the full Board OED's work program and budget, and makes proposals for studies.

a. Modus operandi

The guiding philosophy is that evaluation activities should be comprehensive (all operations should be covered); objective and fair (the evaluation process protected from censorship); transparent (reports should be made available to governments, management, and staff of the Bank, and the process reviewed by the JAC); and participatory, (the operational staff have the responsibility for self-evaluation, to be monitored and assessed by the evaluation staff; the views of the borrower should also be taken into account).

At the completion of a project (or program), the responsible operational unit prepares a Project Completion Report (PCR) appraising the project in terms of its original objectives, implementation, compliance, efficiency, social and economic impact, and a number of other considerations. The OED reviews each PCR to ascertain its completeness and adherence to the Bank's formal requirements.

All adjustment and policy-based programs, as well as about 40 percent of projects, are the subject of an OED audit, which examines documentary

---

1/ Appendices I and II provide a fuller account of the evaluation practices and experience of the World Bank and other institutions, respectively.

material, interviews operational staff, and (in most cases) visits the borrowing country for on-site discussions. The audit report, incorporating staff and national authority comments, and including a section on findings and lessons for the future, is presented to the Board, management, and senior staff as a Project Performance Audit Report (PPAR).

For projects, the audit compares the objectives of the operation (including their appropriateness) as originally stated (at the time of presentation and Board approval) with the outcome at the completion of disbursements. Goals are generally formulated in terms of inputs (e.g., physical inputs); outputs (e.g., roads to be completed); and outcome (e.g., changes in transport costs). The OED uses various qualitative and quantitative indicators in its assessment including the economic rate of return, the cost/benefit ratio, the sustainability of benefits, institution building, and cost overruns.

Structural and sectoral adjustment programs are assessed in the broader context of the country's macroeconomic or sectoral conditions, its institutional structure and capacity, the available policy options, and the global policy environment. The task of evaluation is to determine whether the policies supported by Bank resources were the appropriate ones in the circumstances, the extent to which they were implemented, and how successful they were in reaching their objectives.

Using the PCRs and PPARs as the basic inputs, the OED prepares, with the aid of additional research, many other reports and studies that draw broader lessons regarding the Bank's policies, programs, and activities. These include the Annual Report (a summary of activities); Evaluation Results (an annual summary of evaluation findings of the previous year); Country Reviews (offering a broad sweep of the Bank's relations with the subject country over a one-to-three decade period); Country-specific Sectoral Reviews (in-depth reviews of particular sectors in individual countries); Impact Evaluation Reports (a second look at operations some 5-10 years after completion); Operational Policy Reviews (examining how well the Bank applies its policies, as well as the efficiency and effectiveness of these policies in reaching their intended goals); and Sector Studies (covering matters such as project selection and design and often examining the Bank's experience in a group of countries).

b. Dissemination and use of findings

The above-mentioned reports and studies are distributed to the Board, management, and staff of the Bank, as well as to member countries; some are published. To improve their impact, OED reports carry evaluation summaries focusing on the major aspects of the operation in question, including lessons for the future. Also, OED publishes the Précis, a newsletter that presents the findings of major audits and studies. Another medium of dissemination is the computerized evaluation database which provides the operational staff with user-friendly access to findings and lessons of Bank experience. An essential channel to ensure that evaluation findings reach their primary "consumers" is person-to-person discussion ranging from informal contacts to more formal occasions such as seminars, briefing sessions, and retreats.

The Bank actively encourages the absorption of evaluation findings by the operational staff and their incorporation into future operations. An operational directive, following the findings of a task force on Dissemination and Utilization in 1989, provides that "managers and staff should seek out the findings of relevant OED reports, disseminate them to others concerned, and ensure that they are taken into account in future activities." Relevant OED findings are to be cited in reports supporting or proposing Bank operations; these findings must be consulted early in the project cycle, and the Project Brief (or draft Appraisal Report) must identify the OED document consulted as well as indicate how the key findings and recommendations have been taken into account.

The "management responses" provide another method of formalized feedback; Bank managers prepare responses assessing the validity, significance, and implications of evaluation findings, and identify a plan of action for their use. Also, the Annual Report on Implementation and Supervision prepared by the Central Operations Department presents a summary of initiatives on dissemination and feedback and affords an opportunity for OED staff to comment on work being implemented.

c. Staffing

OED's budget for professional staff for FY 1992 envisages 60.5 staff years (including 17.2 consultant years) for a total cost of $7.9 million. Included are 44 higher-level and 23 support staff positions. OED has also accepted support for specific audits and studies of mutual interest from other donor agencies, the largest of which has been the Canadian International Development Agency (CIDA). In FY 1992, OED expects to review 296 PCRs and to audit 137 of them. In addition 28 studies are being prepared, 12 of which will be completed in FY 1992; these include the Second SAL/SECAL Overview. 1/

d. Conclusions

It is evident that the OED is helpful in enhancing the Bank's internal and external accountability and transparency, in facilitating feedback from the lessons of past experience, and in providing a systematic, quantified, and accessible track record of the institution. Nevertheless, appraising the effectiveness of the Bank's evaluation procedures is not easy and any conclusions must remain largely judgmental.

1/ A first report was prepared in 1986 covering 15 of the early structural adjustment loans; the present study is based on a broader sample of about 100 SALs/SECALs for which PCRs or audits have been completed.

The general consensus among Bank staff with whom the Task Force met is that evaluation as currently conducted makes a useful contribution to operations, and may become more important as new criteria evolve for judging Bank activities. The OED also serves as an important channel for generalizing throughout the operational departments the experience of activities in a particular region or sector. Both the Board and the Bank's management firmly support the incorporation of evaluation findings in operational work, thus promoting productive cooperation between the operational and evaluation staff.

At the same time, Bank staff entertain a number of reservations regarding current evaluation procedures. First, the fairness of analyzing shortcomings many years after the beginning of a project or program, and with the full benefit of hindsight, is questioned by some. Also, evaluation involves ex post appraisals of results against original objectives and these may not have been clearly defined, complicating the task of assessment. Second, the time-lag between implementation and the evaluation process is long (the cycle may take ten or more years); thus evaluation results tend to be dated. Some feel a need for a more dynamic and "relevant" system of evaluation, taking into account changing circumstances. Third, while evaluation findings have become more targeted and issue-oriented, there is room for further progress, including a better integration of OED's work with the evaluation work that is undertaken by the research and policy departments.

2. Other institutions 1/

Broadly speaking, the evaluation systems of other multilateral financial institutions (and also many national aid agencies) are similar in intent and purpose, as well as in many of their procedures, to those of the World Bank. In virtually all cases these systems are separate from the main operational framework of the institutions they serve, and enjoy a considerable independence in reaching and presenting findings. However, practices also differ in some significant ways from those of the World Bank.

Institutionally, the set-up of the Asian Development Bank presents a number of interesting features. The Post Evaluation Office (PEO) of the ADB

1/ The organizations consulted include the African Development Bank (AfDB), the Asian Development Bank (ADB), the European Bank for Reconstruction and Development (EBRD), the Inter-American Development Bank (IDB), the OECD, the United States Agency for International Development (AID), the General Accounting Office of the U.S. Congress, the Overseas Development Administration of the United Kingdom (ODA), the Swedish International Development Authority (SIDA), and the Japanese Overseas Economic Cooperation Fund (OECF). It is interesting to note that the EBRD, which came into being only in 1991, has already established an evaluation unit. This is in contrast to most other institutions where such units were created many years after lending operations began.

reports to the President of the ADB. He forwards PEO's evaluation reports, studies, and findings to a committee of ADB's executive board, the Audit Committee, for their review. The President does not modify the findings of the PEO. Although the PEO is an integral part of the ADB's staff structure, the members of the Audit Committee appear satisfied with its judgmental independence. This, together with the strong support of the ADB's management for the evaluation function, has served to make the current arrangements acceptable to all sides. Indeed, some consider that, for a relatively small institution such as the ADB (with a total professional staff of about 650), these arrangements have certain advantages in that an evaluation unit completely outside the staff structure could run the risk of becoming isolated and ineffective.

The Inter-American Development Bank, on the other hand, has (at present) two evaluation units. One (the Operation Evaluation Office) reports to the Bank's management through the Controller, and the other (the Office of External Review and Evaluation) is answerable directly to the Executive Board. The two units' functions and procedures differ; the former comes closer to the evaluation system of the World Bank.

As in the World Bank, other evaluation units have considerable leeway in deciding which activities to evaluate. In several bilateral agencies, the operational departments may propose particular operations that would be suitable candidates for evaluation. Also, in contrast to World Bank practices, many national aid agencies have a small permanent staff and make extensive use of outside consultants.

Another point of difference concerns the linkage between evaluation and current work. To ensure full independence from current operations, the World Bank's evaluation is deliberately ex post in nature; there is no involvement whatsoever of the evaluation staff until a project or program has been completed. By contrast, in both the ADB and IDB, the evaluation units comment at the various stages of the loan approval process; in this way, they can directly influence the incorporation of past evaluation findings into current operations.

Staff views regarding the benefits of evaluation findings vary considerably in the institutions surveyed: some welcome (or, after initial skepticism, have come to welcome) evaluation, believing that an independent and systematic way of making operational staff aware of the lessons of the past is useful and can lead to improved project/program design and implementation. Also, evaluation can help pinpoint deficiencies and problems. At the same time, others are resistant and defensive about evaluation. There are reservations about individual evaluation findings (in particular that they are often dated, uninformed, and perhaps simplistic). Findings bearing on broader and longer-term issues, including sectoral summaries and studies that provide a deeper understanding of success or failure of programs and projects, are found to be of greater interest. The lack of sufficient time on the part of operational staff to absorb the output of the evaluation units is noted by many.

V. Characteristics of an Evaluation Office in the Fund

In considering the establishment of an Evaluation Office in the Fund, a number of fundamental issues must be addressed at the outset. These include the organizational status of the Office; the scope of its work and how it would go about performing it; arrangements for the dissemination, discussion, and absorption of its findings; and the staffing requirements of the Office, including its Director.

While clear understandings on certain key issues are essential from the beginning, definitive decisions are not necessarily required on all aspects of the Evaluation Office. The initial terms of reference and guidelines on operating procedures can be sufficiently broad to allow the Office to begin its work while also allowing detailed arrangements to be formulated and/or modified in the light of evolving experience.

1. Organizational status

The issues that are basic in defining the organizational status of an Evaluation Office within the Fund are: arrangements for submission and clearance of its findings; the appointment of the Office's Director; and the determination of the Office's work program. These factors will determine to whom the Office is effectively answerable in its work and the extent of its independence in reaching its findings, and hence, to a large extent, its credibility.

In principle, responsibility for these matters could lie either with the Managing Director or with the Executive Board, and some of the merits and drawbacks of the two options are discussed below. Important as this issue is, it must be seen in perspective. Under either option, and in line with the traditions of the Fund, consultation between the Managing Director and Executive Directors would naturally be expected to precede all major decisions concerning an Evaluation Office.

Whether findings of the Evaluation Office are submitted directly to the Executive Board or, in the first place, to the Managing Director and then forwarded to the Board, would be an issue of secondary importance as long as (i) the Director of the Office has the full support and trust of the Board and the Managing Director, and (ii) the Office enjoys independence--from the Executive Board, the Managing Director, staff, and member countries--in reaching its findings and has a major input in the selection of its work program. To ensure such independence, findings would need to be the sole responsibility of the Director of the Office, subject only to the condition that any significant differences of opinion with the main interested parties who will have seen reports in draft (i.e., the Managing Director, senior staff, and the relevant national authorities) be accurately recorded. Equally, to be entrusted with such discretion, it would be of paramount importance that the Director have the necessary professional and personal qualities. (See subsection 7, below.)

The arguments in favor of an Evaluation Office reporting directly to the Executive Board are, first, that this arrangement would institutionalize the assurance that the Office would be an investigative arm of the Board and that management and staff would not exert undue influence on the Office's findings; and secondly, that the Executive Board would have the power of final decision over matters relating to the staffing, work program, and other aspects of the running of the Office.

On the other hand, there are reasons to believe that efficiency and effectiveness would be enhanced by having an Evaluation Office integrated into the Fund's regular management structure. First, it would be a natural extension of the responsibilities of the Managing Director for the running of the institution as a whole that he have the right of decision on major matters concerning the Evaluation Office. He--as head of the staff as well as Chairman of the Executive Board--is best placed to ensure that the findings of the Office are incorporated in staff work.

Second, an Office--especially what would be a relatively small Office--reporting to the Managing Director would run a lesser risk of becoming isolated from the rest of the Fund. Cooperation and productive dialogue between the Evaluation Office and the Fund staff at large--crucial for the Office to be able to perform its functions in an effective manner--would be easier if the Office were located within the staff structure, reporting to the Managing Director, rather than being attached in a sui generis fashion to the Executive Board. With the Evaluation Office and the rest of the staff both reporting to the Managing Director, the risk of unnecessary confrontation between the two would also be lessened. Similarly, adequate staffing of the Office and staff mobility between the Evaluation Office and the rest of the staff--both, as later explained, important to the success of the Office--would be facilitated if the organizational status of the Office were on a par with that of other Fund departments.

To meet the genuine concerns about the independence and credibility of the Evaluation Office while avoiding some of the potential drawbacks mentioned above, the following arrangements could be considered:

- The Evaluation Office would be an autonomous unit (not attached to any other Fund department) reporting to the Managing Director.

- The Executive Board would establish a special committee (say, the "Evaluation Committee") to oversee the activities of the Evaluation Office in the sense of giving detailed consideration to its work program and findings and to the appointment of its Director (see below). Such a committee would report to the full Executive Board as necessary.

- The Managing Director would consult Executive Directors informally prior to making a nomination for the Head of the Evaluation Office; the agreed nominee would subsequently be formally approved by the Executive Board. (This procedure would go beyond Rule N-12 which requires the Managing Director only to inform the Board of intended appointments.)

- The total staff complement of an Evaluation Office would be determined as part of the Fund's budget and would be approved in the normal way by the Executive Board following a proposal from the Managing Director.

- Management would submit the Evaluation Office's work program for consideration by the Evaluation Committee and by the Executive Board. The Board would also discuss the annual report of the Evaluation Office.

- The Evaluation Office would submit its findings to the Managing Director who would transmit them to the Board. The Evaluation Office would ensure that any significant differences of opinion--on the part of management, staff, or national authorities--would be clearly reflected in its reports. (While management would have the option to have dissenting views on its part incorporated in the reports, absence of recorded comment would not necessarily imply complete agreement to the contents.)

2. Scope of work

The terms of reference of the Evaluation Office could be sufficiently broad to allow for possible investigation of all substantive issues arising in the discharge by the Fund of its obligations to the membership. However, at least in the early stages of its existence, the emphasis of the Office's work should be, first, on evaluation of programs supported by Fund arrangements. 1/ The Fund's relations with member countries that are actual (or prospective) users of its resources constitute a vital part of its overall activities and the one of most interest to many of the proponents of an Evaluation Office. Issues worthy of evaluation could also arise in connection with the Fund's relations with countries with whom extensive negotiations have been held but with which, in the end, no program was agreed. A second priority area is the Fund's technical assistance, which is increasingly important in Fund operations. Financial audits and matters of internal organization, systems, and administration would not come within the purview of the Evaluation Office but would continue to be handled by the Office of Internal Audit and Review.

The suggested initial emphasis of the Evaluation Office on Fund programs and technical assistance would not exclude other matters from the

---

1/ Including programs under stand-by and extended arrangements and the SAF and ESAF. In the case of multiyear arrangements, it is suggested that evaluations should normally cover the whole period of the arrangement rather than its component annual programs. Thus the word 'program' and 'arrangement' are in effect used interchangeably below.

Office's potential purview, especially once it had gained experience and maturity. In particular, it is important not to lose sight of the central importance of the Fund's surveillance responsibilities. These arise in relation to all member countries, irrespective of their use of Fund resources at a particular time. Hence it would be important to ensure that the Evaluation Office's terms of reference would allow it to investigate all pertinent matters arising in the context of surveillance.

In discharging its mandate, the Fund undertakes a wide range of other activities reflecting important facets of its relations with its member countries, many of which are complementary to the aims of adjustment programs supported by use of Fund resources (e.g., research, the collection of statistics, and training). While these too might benefit from evaluation, resource constraints and the need to set priorities suggest that they should not be among the initial list of topics to be covered by an Evaluation Office. Furthermore, some of these topics might be appropriate for review by special panels.

In considering the implications of establishing an Evaluation Office for the redistribution of work assignments within the Fund, a number of important points should be kept in mind. First, it must be repeated that the evaluation function in the Fund would not--and should not--become the sole responsibility of the Evaluation Office; self-appraisal as a routine function should be encouraged in all departments. Second, it would be expected that PDR would, for the foreseeable future, continue to have responsibility for the general conditionality reviews; however, as the Evaluation Office becomes established, its findings could increasingly make a contribution to such reviews. Third, general studies on specific aspects and effects of Fund policies would, it is expected, continue to be prepared by the staff of other departments, as well as by the Evaluation Office.

3. Evaluating Fund-supported programs

It would be inappropriate--and perhaps impossible--at the outset to prescribe a standard methodology for evaluation of Fund programs and of program-related discussions with national authorities. 1/ Institutions involved in project lending have established certain quantifiable criteria for judging success and failure. However, the establishment of such unambiguous criteria would be considerably more problematic in matters concerning the selection of the goals and instruments of macroeconomic policy. Such issues are, to a large extent, judgmental and the choices to be made are--quite legitimately--affected by social and political preferences. Moreover, the necessary empirical evidence is often incomplete, or does not point to unambiguous conclusions, thus complicating the task of evaluation and emphasizing the need for judgment.

1/ For evaluation of other topics (e.g., those related to surveillance) the appropriate method would depend entirely on the precise issue in question.

Even though standard methodologies cannot be prescribed, an illustrative list of topics that could arise in evaluating Fund programs is set forth below. For convenience, the topics are presented in the form of questions. The choice of which questions to address would be for the evaluators to make on a case-by-case basis and, it must be stressed, there is no suggestion that they follow a "checklist" approach. In the case of any individual evaluation, many of the questions listed may not need to be addressed.

"Pre-program" relations

- Was the Fund's advice in the "pre-program" period appropriate? Were emerging problems satisfactorily identified?

Program negotiation and approval

- Did the program start at a suitable time? Could a program have been agreed earlier (perhaps when economic problems were less acute)? Alternatively, should agreement on a program have been postponed to a later date (perhaps when it might have been possible to agree a stronger program)?

- Were adequate steps taken to ensure that the Fund's previous experience in the same country, or in other countries in similar circumstances, was adequately taken into account?

- Were the Fund and the national authorities sufficiently aware of each other's views on the basic features of the program's design? Was the program as negotiated one to which the authorities were fully committed and which they believed they could implement?

Program design--objectives

- Were the objectives of the program for the major macro-economic variables appropriately selected? Were they consistent and clearly specified? How were choices between competing objectives handled? Did they realistically take account of political and social constraints and of the implementing capacities of the country's institutions?

- Were the program's assumptions regarding the external environment (e.g., terms of trade, terms and availability of external financing) realistic?

Program design--policy instruments

- Were the policy instruments--including "prior actions"--well chosen to achieve the program's objectives? Did they take proper account of the characteristics of the economy of country in question? How were choices between alternative instruments handled?

- Were the selection, specification, and number of variables to be subject to performance criteria appropriate? Were satisfactory arrangements made and followed for monitoring and reporting to the Fund?

- Was technical assistance employed to facilitate policy implementation? If not, could it have been? If it was, was it successful?

- Were the structural and institution-building policies included in the program appropriate?

Program design--Fund financing issues

- Was the total amount and phasing of purchases under the program appropriate in light of the country's balance of payments need, the Fund's access policy, and the history of its previous financial relations with the Fund?

- What criteria were applied in assessing the country's future ability to make repurchases? Were the medium-term scenarios used realistic?

Relations with the World Bank and other financing sources

- What steps were taken to coordinate policy actions in the context of the Fund program with the advice and activities of the World Bank (and other multilateral institutions)? Were these arrangements satisfactory for the country, the Fund, and the Bank?

- How were relations with official and private external creditors reflected in designing and negotiating the program? How were understandings reached and coordinated regarding the amounts of new financing, debt rescheduling, and debt and debt-service reduction that each of the parties would provide? Were arrangements satisfactory to the various parties involved?

Role of the Executive Board

- What was the contribution of Executive Directors--notably through Board discussions--to the design, negotiation, and (where relevant) modification of the program? Were any particular issues or concerns focused upon by the Board?

Program implementation and impact

- Were policies agreed under the program implemented as originally envisaged? If not, what were the reasons?
- Were unexpected developments during the program adequately handled in mid-term reviews or by other means? Was the rationale for any waivers granted or modifications to the program appropriate?
- Were the policies followed under the program sustained after the expiration of the arrangement with the Fund? What are the prospects for their sustainability in the longer term?
- What was the relationship between the extent to which agreed *policies were implemented and performance criteria observed*, on the one hand, and the degree to which the targets of the program were achieved, on the other?
- What was the overall effect of the program on the country's short-term economic performance and its longer term prospects?

4. Evaluating technical assistance

In broad terms, the topics for potential evaluation would include: (i) the identification of technical assistance needs; (ii) the subject matter of the advice given; (iii) the manner in which assistance is provided; (iv) the preparedness and willingness of the host country to absorb and act upon the advice provided; (v) the cost effectiveness of assistance in developing expertise in the host country and in contributing to economic performance in general, and to performance under Fund programs. Many technical assistance assignments involve only a small number of individuals and their impact can be affected by a wide range of factors specific to the particular task in question. Hence, there may often be an advantage in the Evaluation Office considering simultaneously the experience of broad groups of countries facing similar problems (e.g., banking supervision procedures in transition economies or tax administration in a particular region).

At some stage, the findings in regard to technical assistance might help in Board decisions on the total volume of resources which it is appropriate for the Fund to devote to technical assistance and the manner in which that total should be allocated across countries and topics.

5. Preparation, discussion, and dissemination of findings

a. Selection and timing

The number of programs (arrangements) that it would be possible for the Evaluation Office to review would depend in part on the size of its staff. To evaluate all programs would require a large staff except at the risk of considerable superficiality. In the early stages of its work, it would appear appropriate for the Office to place the emphasis on evaluating some programs in depth rather than to attempt to cover a large proportion of all programs. Selection of appropriate programs would be crucial; those programs could be chosen whose lessons are likely to be of broad interest and relevance, either because performance has been notably successful or because the experience failed in major respects to come up to expectations. It would also be appropriate, to the extent possible, to aim for representative coverage--in geographical terms, as between low- and middle-income countries, as between "successes" and "failures", and as between prolonged users of Fund resources and others. In present circumstances, it would also be especially appropriate to evaluate as thoroughly as possible programs with countries undergoing transition from a planned to a market economy.

As to timing, evaluation should be undertaken sufficiently soon after the conclusion of a program to allow any lessons for the future to be drawn before they are overtaken by events. A relatively short lag between program completion and evaluation would also have the advantage of increasing the chances that personnel--among the Fund staff and in the country--would still be occupying the same positions as during the program period and have events of the period fresh in their memories. On the other hand, it would also be important that relevant data be available to the evaluators so that they could make an at least preliminary assessment of the sustainability of policies and of the performance of the economy in the post-program period. Weighing these considerations, evaluations could begin around 12-18 months after the expiration of the program to be evaluated.

b. Information requirements

The staff of the Evaluation Office would have access to all relevant documents prepared for the Executive Board and to the minutes of Executive Board meetings. They would also have access to all pertinent information held by the staff. The concerned departments of the Fund would be expected to make available to the Evaluation Office mission briefing papers, back-to-office reports, minutes of meetings, and other memoranda and documents. As employees of the Fund, the staff of an Evaluation Office would be bound by the same rules on confidentiality as other staff members.

The Evaluation Office would be bound to use confidential information responsibly. Confidential information would not be included in any published reports.

As indicated earlier, self-appraisal by all departments would be important. Ideally, such evaluations of programs would be prepared shortly after the end of the program (say, within six months) so as to be available to the evaluators when they begin their work.

The Evaluation Office would be expected to be in regular informal contact with the staff at various stages in the preparation of its reports. The Office would also obtain the views of national authorities and, where necessary, visit the country concerned for discussions. Finally, the Office could use the expertise of outside parties (for example, academic commentators) with knowledge and experience relevant to specific topics.

c. Reporting

In line with the proposals above, the main output of the Office, especially at the start, would be reports on individual country programs and technical assistance. As the number of individual programs evaluated increased over time, it might be appropriate for the Evaluation Office to produce reports on general themes synthesizing the experience of various countries as well as reports surveying the experience of a particular country over an extended period. In addition, as noted earlier, it could prepare studies on issues that are thematic rather than being addressed to individual country experiences, on topics arising in connection with surveillance and, possibly, other Fund activities. The Office would also prepare an annual report of its work and the principal evaluation issues that had arisen.

The Evaluation Office should develop appropriate procedures for circulating its reports in draft to all interested parties for comment and incorporate changes it agreed with. As earlier noted, any remaining, significant differences would be duly noted in final reports.

The natural forum for discussion of final reports would be the proposed committee of the Executive Board charged with oversight responsibility of the Evaluation Office. In addition to members of the Committee, staff from the Evaluation Office and the relevant departments of the Fund could be present for a "seminar-type" discussion of the report. Minutes of such discussion could be kept. Consideration could also be given to circulation among the staff of short summaries of points of general interest to emerge from the discussion of any particular report.

Evaluation reports on individual programs (or technical assistance) would not be published. However, an annual report summarizing the activities and major findings of the Office and any thematic studies could be considered for publication. It would also be helpful, in disseminating its findings, for the Evaluation Office to develop a computerized system to which all Fund staff and the Executive Board could have access.

6. Absorption

Even the best evaluation findings would be of little use if they were not absorbed by the operational staff and incorporated into current and future work. The Evaluation Office's findings would be disseminated among the staff, possibly together with summaries of the Executive Board's discussion of them, as discussed above. However, it would be for management and senior staff to institute arrangements to ensure that operational staff took due account of the Office's findings. Some institutions (such as the World Bank) have issued formal directives to this effect. However, at least equally important would be cooperative relations between the Evaluation Office and the operational staff; informal and spontaneous relations are likely to be of more value than formal contacts undertaken to comply with directives. Close working relations would, to an important degree, depend on the operational staff being convinced of the positive contribution of evaluation findings to their work; this, in turn, would be affected by the objectivity, quality, and credibility of evaluation findings.

7. Staffing requirements

It would be an error to assume that evaluation can be had "on the cheap": an adequate complement of high quality staff is essential if an Evaluation Office is to be effective. Indeed, experience of other institutions suggests that the provision of appropriate staffing is more important, for instance, than to whom the Evaluation Office formally reports. A highly qualified--and impartial--Director and staff would be particularly likely to enlist the confidence of the Executive Board, management, and staff. Further, an Evaluation Office thus constituted could better establish close, professional relations with the staff at large, and create the conditions for a candid dialogue and exchange of information, factors that would be of paramount importance in its work.

a. Post of Director

The Director of the Evaluation Office should be a person of proven competence, held in high regard by his peers, and of independent and impartial judgment. Furthermore, he/she should not stand to gain or lose professionally from the evaluation findings of the Office, including (where warranted) critical conclusions regarding Fund programs and policy recommendations. In other words, the Director should not be placed in situations of actual or perceived conflict of interest. It is also essential that the Director be familiar with the range of problems and situations that the Fund typically faces in its operational activities, and be knowledgeable of the Fund's policies on broader issues and concerns.

One possibility would be to appoint as Director a senior staff member, with a successful career, who had reached the age where he was expected to retire within the next five years. In that case, an appointment could be made for up to, say, a five-year period, after which he would separate from the Fund. In selecting the Director (as well as the staff) of the

Evaluation Office, due consideration should be given to suitably qualified persons who have left the Fund.

Alternatively, the Director could be appointed from outside the Fund staff. In such a case, the incumbent should also be closely familiar with the policies, activities, and operations of the Fund. For instance, he could have had an eminent career in government or finance. Again, the appointment should be for a fixed term.

In either case, it is evident that the Director would have a deciding influence on the general ethos of the Evaluation Office, its credibility and effectiveness, and its acceptance by the Executive Board, management, and staff. Hence, the greatest care would need to be exercised in making the appointment.

b. Staff

As in the case of the Director, it would be important that the staff members of the Office be of high caliber; possess solid backgrounds in macroeconomics and finance, and good analytical judgment; and have an ability to cooperate with the staff at large. A period of operational experience and, in particular, familiarity with operational issues typically faced by the Fund, would be a distinct advantage. For these reasons, the Office's staff would probably have to be largely recruited from current and former Fund staff members. It is also assumed that the bulk of the evaluators would be at senior economist level. At the same time, there would be advantages to having a portion of the staff--whether on a short term or longer-term basis--recruited directly from outside, provided they possess the necessary qualifications. Also, the participation of an experienced member of the Bank's Operation Evaluation Department might be helpful to the early work of an Evaluation Office.

Mobility between the staff of the Evaluation Office and the mainstream staff of the Fund (in particular, area departments) would contribute to maintaining an alert and informed Office. For this purpose, the Fund's personnel policies would need to emphasize the professional rewards to staff of spending a period of time in the Evaluation Office; secondment to the Evaluation Office should be recognized as a positive element in one's career. The experience of other institutions suggests that if mobility is not encouraged and actively promoted, the Evaluation Office could run the risk of becoming a preserve--if not a pre-retirement pasture--for staff members who (for whatever reasons) have run into career obstacles in other departments. A competent staff and mobility go hand in hand; only if the Office were staffed with persons of high caliber would it be effective and, in turn, attractive as a positive career step.

c. Total staffing requirement

The number of staff required for the Evaluation Office would largely depend on (i) the nature and functions of the Evaluation Office that is eventually established and (ii) the expected coverage of the Office's activities in each instance.

Consistent with the suggestions made in previous sections, one could envisage the work program of a fully staffed Evaluation Office as encompassing (i) the evaluation of half of all Fund programs that ended in the preceding year, (ii) evaluation of selected aspects of the technical assistance provided by the Fund, (iii) the preparation of a few thematic studies, and (iv) active participation in the dissemination of its findings. Drawing on the experience of the World Bank and other institutions, it may be assumed that a minimum of 12 staff-weeks would be required for each country-program evaluation. If it is assumed that some 30 Fund programs expire in a given year (as will be the case in 1992), this would total 180 staff-weeks for evaluating programs. One could also envisage that two persons would be occupied with evaluation of technical assistance in non-program countries for a total of 90 weeks worked. In addition, adequate provision would need to be made for general studies and for dialogue with the operational staff to discuss current evaluation work and past findings. One may (conservatively) assume a further 90 staff-weeks for these purposes.

The total staff requirement, based on the above assumptions, would amount to 360 staff-weeks. Dividing this figure by 45 (which is the length of one staff year after allowance for vacations and holidays) yields approximately 8 staff-years. To this must be added the post of the Director and at least one research assistant, bringing the complement of professional staff to, say, 10 individuals. With support staff added, one may conservatively envisage an overall staff requirement of, say, 13 persons. 1/ Of course, the Evaluation Office would not be established in its full complement from the beginning; it would take time for its activities to be properly defined and for the Office to be fully staffed.

VI. Conclusions and Recommendations

1. By offering independent and disinterested assessments of the Fund's activities, an Evaluation Office, separate from the mainstream of the Fund, could provide a useful focal point for drawing the lessons of past experience. It could lend greater acuity to analyses of policies and practices, thereby contributing to better program design and implementation, and enhancing institutional memory.

2. The potential benefits would not flow automatically from the establishment of an Evaluation Office. The effectiveness and credibility of the Evaluation Office would hinge on many factors, including most notably:

1/ It may be noted that the World Bank's OED employs some 60 professional staff and the Asian Development Bank has a complement of 25 professional staff. The bilateral aid agencies have much smaller staffs but make extensive use of consultants.

(a) the firm support of the Board and management for the evaluation function, and the cooperation of the operational staff; (b) the independence of the Office from day-to-day activities, including policy formulation and program implementation; (c) the ability of the Office to issue its findings in a candid, transparent manner, without pressure or censorship; and (d) the appointment as Director of a person of recognized status, at once closely familiar with Fund activities and possessing first class analytical abilities, and supported by adequate personnel of the highest quality. Without these features, evaluation findings could well be ignored or treated as a formality. At the same time, it must be stressed that the drawing of clear and practically applicable lessons may be difficult for analytical and other reasons. Moreover, there could be certain drawbacks to the establishment of an Evaluation Office including resistance from the staff and excessive caution in carrying out its duties.

3. The Fund currently engages in a variety of appraisal activities. Area departments routinely assess recent programs in Article IV consultation reports and in the papers supporting successor programs. These assessments, however, only appraise the performance of countries, not that of the Fund, and they are prepared by the department largely responsible for the program reviewed. Periodic reviews, called for under the conditionality guidelines, and the comparable reviews of SAF and ESAF experience, bring together much valuable material on the experience with programs. But the focus of these exercises is heavily on comparing outturns (for growth, inflation and the balance of payments) with forecasts made at the beginning of program years--often referred to as "targets"--and, for SAF/ESAF, on operational issues. While these exercises contain, in varying degrees, building blocks useful for the evaluation of the underlying programs, they fall well short of constituting the kind of incisive, independent, evaluation of its programs that the Fund should have. Various functional departments from time to time prepare studies analyzing general or particular economic features of Fund programs; such studies should continue side by side with the activities of an Evaluation Office.

4. A review of the evaluation practices and experience of other multilateral financial institutions, as well as of a number of national aid agencies, shows that the evaluation function is regarded as an important way of enhancing internal and external credibility and facilitating the absorption of the lessons of past experience. In most cases, evaluation units have the strong support of their respective governing body and management (whether the unit reports to the former or the latter). Discussions with the operational staff of the institutions surveyed--the so-called "consumers" of evaluation reports--suggest that, by and large, the evaluation function is recognized as an important element in the institutions' activities even where, initially, it had met with resistance and skepticism. In general, cooperation between the operational and evaluation staff is good and in some instances special arrangements have been made to ensure that evaluation findings are incorporated in proposals for new projects or programs. At the same time, there are certain significant differences between the institutions in question and the Fund.

For instance, most of the institutions are still largely project-oriented in their work; hence the approach to evaluation and analysis of past experience is in many ways more straightforward and amenable to quantification than the macroeconomic context that characterizes Fund work.

5. In contemplating the establishment of an Evaluation Office in the Fund due consideration must be given at the outset to the principal features that would comprise an adequate evaluation function. Many of these are interrelated; others present a choice; and others still can only be refined and determined in the light of evolution and experience. The principal aspects of an Evaluation Office are set forth below, in points 6 through 12.

6. The Evaluation Office, as a separate unit, could report directly to the Executive Board, or, as is the case with existing Fund departments, to the Managing Director. While the institutional location of the Office is clearly an important issue, the more significant point is that it should have objectivity, independence of judgment, and an unfettered ability to reach and to present its findings. Also, for a relatively small institution like the Fund, it would be important to ensure that the Office (which presumably would also be relatively small) did not become isolated and lose touch with the operational staff, a risk that could attend placing the Office outside the regular staff structure. One way of assuring functional independence, yet keeping the Office within the traditions and structure of the Fund, would be for it to report to the Managing Director while enjoying a clearly recognized right to present its findings to the Board without amendment. Under such an arrangement, the Managing Director would have the option to have his views, where they differed from those of the Evaluation Office, recorded in the evaluation reports of the Office. At the same time, a special committee of the Executive Board could exercise a supervisory role over the Office's activities, including the appointment of the Office's Director, review of its work program, and discussion of evaluation reports and findings; as necessary, such a committee could make recommendations to the full Executive Board for final decision.

7. The prime (but not exclusive) focus of the Office's work would be to evaluate Fund operations, that is, Fund programs and technical assistance and, where appropriate, the Fund's advice in the context of Article IV surveillance with active or prospective program countries. In addition, and within the limits of its resources, the Office could undertake studies of other aspects of Fund relations with members. After a sufficient number of program evaluations have been completed, they would clearly provide an important information base for broader cross-country analyses that the Office might be requested to undertake. The initially rather narrow definition of the scope of the Office's activities is motivated by the need to concentrate the activities of a relatively small Office on the most urgent tasks of evaluation.

8. The Office should be assured of free access to all pertinent information and data. The operational staff should be required to cooperate with the evaluation staff. There could be some resistance to such

cooperation on the part of the operational staff but "creative tension" is not always detrimental. The evaluation staff should also have the possibility of country visits for first-hand discussions with officials.

9. In evaluating Fund programs, the Office would be expected to look into all elements that are relevant to economic performance under such programs. These include: (a) the economic framework in which a program was formulated, i.e., the underlying economic situation and circumstances, the magnitude and nature of problems, the size of the financing gap, etc.; (b) program design, including major targets and recommendations, as well as the transmission and response mechanism assumed by the program; (c) major assumptions underlying the program, including assumptions regarding external economic developments and the technical and political implementation capacity of the authorities; (d) length of program and amount of financing provided under the program, and assumed to be generated as a result of the program; (e) exogenous developments and their impact on program performance; and (f) implementation of the program. No single, specific methodology can be recommended in undertaking the above analysis and no simple measure of "success" or "failure" would be appropriate. The nature of Fund activities--in that the Fund deals with macroeconomic issues that are complex and multifaceted and have a high political content--as well as the broad character of certain performance goals (such as "viability" in the balance of payments), would make it difficult to develop such criteria. With respect to technical assistance, the Office would investigate the effectiveness of the assistance provided, in particular its contribution to economic performance.

10. The findings of the Evaluation Office should be as up-to-date and targeted as possible; they should always focus on the activity being evaluated and not pass judgment on staff members. On timeliness, the Fund enjoys an advantage over other institutions in that Fund programs are of relatively short duration and thus may permit findings that are both more current in context and more "actionable" in practice. The evaluation of a program should start 12-18 months after the end of the program period (or arrangement in the case of multiyear arrangements). In addition, evaluation of the longer-term impact of selected country arrangements with the Fund could be undertaken at a later date. Evaluation reports should be concise and "reader-friendly." They should be submitted in draft to the operational staff involved, management, and national authorities; each would have their views on important issues recorded if these differed significantly from those of the Evaluation Office. Final reports should be distributed to the Board, management, and senior staff, as well as to officials of the program country. Supplemental means of disseminating findings should also be developed, including seminars between the evaluation and the operational staff, workshops, and individual contacts. An annual report summarizing the activities and major findings of the Evaluation Office should be prepared for discussion by the Executive Board, and be published.

11. To be effective, evaluation findings must be absorbed into current programs and activities. The active support of management and senior staff would be crucial in achieving this. Further, a requirement could be

instituted that findings on countries be explicitly taken into account in the design of new programs for the same country.

12. It is most important that the Evaluation Office be endowed with adequate staff resources, in number and quality. The amount of resources would largely be dictated by the nature of the activities mandated to the Office and by the expected comprehensiveness of its work. The Director (as a person of recognized competence, highly regarded, of objective judgment, and intimately familiar with Fund operations) could be a senior staff member appointed for a fixed term who would not return to the staff at the end of his or her term. A former member of the staff or Executive Board or a qualified outsider could also be considered. The personnel of the Office should be of high quality and it is desirable that they should have had experience in operational work. Further, service in the Evaluation Office for a fixed period of time should be regarded as a positive element in career development and should be actively promoted by management. This would establish a healthy mobility between the operational and the evaluation staff, to the benefit of both.

The World Bank's Evaluation Procedures and Experience

The Bank has in place an independent, extensive, and systematic evaluation process as an integral part of its activities in general, and program management in particular. Evolved over a period of more than two decades, this evaluation system is generally regarded as the most developed of any financial institution. By imparting greater credibility, promoting accountability, and facilitating feedback, the evaluation process seeks significantly to enhance the Bank's operational effectiveness. This section provides a summary of the salient features of the Bank's evaluation procedures. 1/

1. Genesis and institutional set-up

With some prodding from outside, the Bank's management became convinced in 1970 that an open, independent review process would be of benefit to the Bank for several reasons:

- the assessment of past performance would yield lessons of relevance to current experience and formulation of goals for new operations;

- evaluation would enhance staff accountability and strengthen trust in the Bank; and

- evaluation and accountability would be positive for the image of the Bank.

Beginning as a unit in 1970, an Operations Evaluation Department (OED) was established in 1973 and, in the following year, the post of Director General, Operations Evaluation (DGO) was created at Vice President level.

The **Director General** is independent of the mainstream activities and structures of the Bank; he is appointed by the Board for a fixed term and he is directly responsible to the Board. He retains an administrative link to the President (in that OED's budget is determined by Bank management) and, once his term expires, he may not join (or re-join) the Bank staff.

The DGO's responsibilities are to :

- keep watch over the adequacy of the Bank's overall evaluation functions including OED, and the dissemination of findings and feedback;

1/ This section is based on discussions with the Bank's evaluation staff (who also provided extensive documentation), as well as with selected members of the operations staff. Their cooperation is gratefully acknowledged.

- assess whether the Bank's operations are producing the expected results;

- distill lessons of experience and incorporate them into recommendations designed to increase efficiency of Bank activities and enhance the Bank's responsiveness to member countries' needs and concerns; and

- assist member countries in developing their own evaluation systems.

The **Operations Evaluation Department**'s staff (which is almost exclusively recruited from within the Bank) is also independent of the Bank's mainstream and structure; it reports to the DGO. The OED's principal function is to conduct performance audits on selected completed operations and evaluation studies on specific issues and sectors. Its other functions are to: help disseminate the lessons and findings of evaluations (within the Bank and outside); follow up and assess actions triggered by OED findings and studies; assist in appraising the Bank's overall evaluation system (i.e., including evaluation work undertaken outside OED); and develop and maintain the Bank's institutional memory on evaluation and make it accessible to potential users. Put differently, the OED's ultimate focus is on the Bank's <u>policies</u> and their appropriateness, the Bank's <u>procedures</u> and their integrity, the Bank's <u>processes</u> and their adherence to established rules, and the Bank's <u>effectiveness</u> in promoting lasting development in member countries.

To facilitate its oversight of the Bank's evaluation procedures, the Board has established an eight-member **Joint Audit Committee (JAC)** which, inter alia, oversees and assesses the activities of the OED. In discharging its functions, the JAC considers and comments on OED reports, scrutinizes and recommends to the full Board OED's work program and budget, and may suggest particular subjects for study by the OED.

2. <u>Modus operandi</u>

The general principles guiding the Bank's evaluation activities are that they should be: <u>comprehensive</u> (all operations should be covered); <u>objective and fair</u> (the goals and means of operations should be as explicit as possible so as to allow for later evaluation, and the evaluation process protected from censorship); <u>transparent</u> (reports should be made available to governments, management, and staff of the Bank, and the process reviewed by the JAC); and <u>participatory</u> (the operational staff have the responsibility for self-evaluation, to be monitored and assessed by the evaluation staff; the views of the borrower should also be taken into account).

The evaluation functions carried out by OED 1/ may be viewed as a two-stage procedure consisting of reviewing the self-evaluation undertaken by the operational staff upon completion of a program and project and (in about half of the total cases) doing an independent audit of these projects; and the ensuing impact assessments and other studies that OED undertakes using the evaluation and audit reports mentioned above as basic building blocks.

a. Review and audit

- At the completion of a program or project, 2/ the responsible operational unit prepares a **Project Completion Report (PCR)** which appraises the project in terms of its original objectives, implementation, compliance, efficiency, social and economic impact, and a number of other considerations. 3/ The PCRs, which thus provide the nexus between the operational supervision of a project and its ex post evaluation, are forwarded to the members of the Board 4/ and to OED. The latter reviews each PCR to ascertain its

---

1/ This section concerns the evaluation activities focused on the OED; it is important to bear in mind that the operational, policy, and research staff of the Bank also undertake evaluative work as part of their functions.

2/ Normally within nine months of final disbursement.

3/ Among the questions that the PCR may address are: (i) Were the goals of the operation appropriate and clearly defined? Were these goals realized, in the view of the Bank and the borrower? What were the strong and weak points of the operation? (ii) Are the economic and social benefits of the operations likely to be realized? How will they be distributed among the regions or income groups affected? Were there any unintended economic, social, or environmental effects? When should such a judgment be made? (iii) How well did the operation succeed in relation to its institution-building strategy? (iv) Were the agreed or anticipated policy reforms carried out, and if so how well? (v) Were the arrangements for managing the operation satisfactory? Were the operations's financial objectives, including cost-recovery and self-financing of investment, fulfilled? (vi) Was the operation successfully implemented in its major components? Was it kept within reasonable cost estimates? What changes were made in the course of implementation? (vii) Did the borrower comply with the loan covenants and related agreements? (viii) Could the operation be replicated: could similar operations be more quickly or economically be carried out in the future without risk to quality? (ix) Does the investment remain worthwhile, taking into account all the shortcomings, delays, etc.? (x) How would one assess the World Bank's contribution, from the borrower's viewpoint, as well as in terms of diagnosis, definition of goals, design, and implementation?

4/ Beginning in June, 1992; hitherto PCRs were sent to the Board by the DGO.

completeness and adherence to the Bank's formal requirements. The PCR is also made available to the operational staff and to the borrowing country.

- All adjustment and policy-based programs, as well as about 40 percent of projects, are audited by OED. In so doing, OED examines all documentary material (including minutes and files on the operation), interviews operational staff, and in most cases visits the borrowing country for on-site discussions. After comments from the operational staff and the borrowing country, the audit report, together with the PCR, is presented to the Board, management, and senior staff in the form of a **Project Performance Audit Report (PPAR)**. This report, in addition to assessing the objectives, design, implementation, monitoring and other aspects of the program or project in question, includes a generally tersely worded section on findings and lessons for the future.

The underlying purpose of the PPAR is to present, in an accurate, fair, and candid manner, and with due regard to the views of the borrowers, beneficiaries, and the operational staff, an evaluation of the operational experience, an analysis of any problems with the operation, and an indication of how things might be done better in the future. While the PPAR may be critical, assigning blame in the light of hindsight is not one of its goals. Matters requiring further study are explained, as are any significant information that has been omitted on grounds of confidentiality.

The methodology of the audit is to compare the objectives of the operation as originally stated (at the time the operation was presented to the Board for approval) with the outcome at the time the Bank completes disbursements. Any modifications to the goals made in the course of implementation are taken into account. Goals are generally formulated in terms of <u>inputs</u>, such as cost and physical inputs; <u>outputs</u>, such as physical construction of roads to be completed or teachers to be trained; and <u>outcome</u>, such as changes in transport costs or increase in literacy. While inputs and outputs can be measured with the aid of data collected from the operation, outcome is more difficult to gauge objectively and the OED staff resort to a number of qualitative and quantitative indicators in making their assessment. Such indicators include the economic rate of return, the cost/benefit ratio, the sustainability of benefits, progress with institution building, and cost overruns.

Adjustment programs (such as structural adjustment lending--or SALs--and sector adjustment lending--or SECALs) which are in support of changes in a member country's economic policies, are assessed in the broader context of the country's macroeconomic or sectoral conditions, its institutional structure and capacity, the available policy options, and the global policy environment (including trade policies, exchange rate movements, external debt, etc.; such areas as the environment and the role of women are also covered). The task of evaluation is to determine whether the policies supported by Bank resources were the appropriate ones in the circumstances, the extent to which they were implemented, and how successful they were in reaching their objectives. In such an evaluation, an important consideration is the extent to which microeconomic strategies and policies mesh with macroeconomic reforms.

b. Reports and studies

Using the PCRs and PPARs as the essential building blocks, OED prepares, with the aid of additional research, a number of other reports and studies that seek to distil and complement its findings and draw broader lessons regarding the Bank's policies, programs, and activities.

Annual Report. A summary account by the DGO of the activities of OED including a report on the dissemination and use of evaluation findings, and an account of staffing and use of resources. This report is discussed by the Board.

Evaluation Results for [year]. This annual review, discussed by the Board and subsequently published, summarizes the evaluation findings of the previous year, as well as an analysis of particular topics that have been the subject of an OED study (see below).

Country Reviews. These studies, of which six have been conducted and a seventh is in preparation, offer a broad sweep of the Bank's relations with the subject country over a one-to-three decade 1/ period. These reports are notable for their detailed perspective and blunt language. 2/

Country-specific Sectoral Reviews. In-depth reviews of particular sectors in individual countries assessing the impact of Bank

---

1/ The report on Senegal covered the period 1960-87; the most recent, on Bangladesh reviewed the experience of the 1980s.

2/ For example: "What is surprising and difficult to understand is how irresponsibly passive the Bank continued to be throughout its policy dialogue with..." "...failure to recognize sufficiently the structural and deep-rooted nature of many of these constraints, such as...repeatedly led the Bank to explain away or excuse poor performance as being caused by cyclical and temporarily exogenous factors." The use of candid language is not limited to the country reviews; OED studies are generally outspoken in their assessment.

The recommendations in the country reports can be general as well as specific. For instance, the Tanzania country report contains 28 recommendations including 9 on policy matters, 7 on the lending program, and 4 on Bank strategy.

lending (for instance, Indonesia's population program, or road maintenance in the Philippines).

Impact Evaluation Reports. These reports provide a second look at operations some five to ten years after completion, and seek to assess the longer-term benefits--and drawback--of the project. These assessments, in many instances, are broader in scope than the original objectives of the projects in question and, in addition to giving greater emphasis to longer-term effects, focus on the impact of projects on their intended beneficiaries.

Operational Policy Reviews. These reports examine how well the Bank applies its policies, as well as the efficiency and effectiveness of these policies in reaching their intended goals. They may cover an individual country or broader policy issues. 1/

Sector Studies. The focus of these studies is country sector work and covers matters such as project selection and design, implementation, sustainability of benefits, etc. Many of these studies examine the Bank's experience in a group of countries (such as, for instance, Bank support for industrialization in newly industrializing countries).

c. Dissemination, absorption, and action

Dissemination

The OED prepares an impressive volume and array of audit reports and studies; indeed, some "consumers" find the menu to be too long. These are circulated for comment at draft stage to operational staff and others within the Bank, as well as to the borrowing member. Upon completion they are presented to the Board, management, and staff of the Bank, as well as to member countries; some are published.

Documentary reporting is the first stage of dissemination and the OED has taken steps to improve its focus and impact. One constant problem is the time availability and absorptive capacity of busy operational staff. To partially meet this problem, OED reports now carry evaluation summaries focusing on the objectives and design of the project or program in question, implementation experience, results, sustainability of benefits, and lessons

---

1/ For example, the 1986 study on structural adjustment lending concluded, inter alia, that quantitative targets for economic outcomes should be avoided as part of conditionality; there was a need for more clearly stated intermediate goals and greater readiness for mid-course corrections; further research and analysis was required on the complex interdependence of policy measures; more rigorous analysis was required of the political sensitivity of individual measures.

for the future. Furthermore, the OED has begun publication in newsletter format of the Précis, an attractive, well written, and accessible presentation of the findings of major audits and studies.

Another medium of dissemination is the computerized evaluation database which carries copies of the summaries of all PCRs and OED reports and studies. The database, which is accessible to all Bank staff through their "All-in-One" network, is complete for the years 1972-91 and is updated twice yearly. For the operational staff member, the database provides ready, user-friendly, access to the findings and lessons of Bank experience of relevance to his or her work.

Important as the above vehicles for dissemination are, an essential medium for ensuring that the findings of the OED reaches their primary "consumers," that is, the operational staff, is person-to-person discussion and dialogue. This ranges from the wide-ranging informal contacts that characterize the cross-fertilization of ideas in any institution, to more formal occasions such as seminars, briefing sessions, and retreats.

Absorption and action

Effective evaluation depends not only on the quality of the evaluative work and its adequate presentation and dissemination, but, crucially, the proper absorption of findings by the operational staff and their incorporation into future operations. The Bank is keenly aware of this phase of the evaluation exercise; it has in place a number of procedures for assuring proper absorption and action and has taken initiatives in recent years to assure that evaluation findings become an integral part of operational work. The main channels for absorption and action are:

- A Task Force on Dissemination and Utilization of OED findings presented its report in March 1989; this was discussed by the JAC and led to the adoption of an operational directive to the Bank staff. It provided that: "Managers and staff should seek out the findings of relevant OED reports, disseminate them to others concerned, and ensure that they are taken into account in future activities. In keeping with existing policies, relevant OED findings should be reflected and cited in Country Strategy Papers, Staff Appraisal Reports, Project Briefs, and President's Reports. It is particularly important that (a) relevant OED reports be consulted early in the project cycle, and (b) the Project Brief (or draft Staff Appraisal Report, if prepared instead) identify the OED document (and management responses, when applicable) that have been consulted, and how the key findings and recommendations have been taken into account."

- The directive cited above generalized and confirmed practices that already existed in some regions. Thus, the pioneering work of the Indonesia Country Department is often singled out as a model (where the then-Director would not approve papers supporting new operations that did not explicitly take into account the findings of OED). The Latin America and Caribbean Vice President has issued explicit directions to his staff requiring papers and reports clearly to reflect the lessons of the past. The Africa region is a keen user of the computerized evaluation database; it has directed that the lessons of evaluation be incorporated in project preparation.

- The Economic Development Institute uses evaluation and audit reports as part of its source material and OED staff participate in some of its sessions and seminars.

- The so-called management responses provide a method of formalized feedback; Bank managers prepare responses assessing the validity, significance, and implications of evaluation findings, and identify a plan of action for their use. Further, the Annual Report on Implementation and Supervision prepared by the Central Operations Department presents a summary of initiatives on dissemination and feedback and affords an opportunity for OED staff to comment on work being implemented.

- Various departments prepare reviews that summarize evaluation findings and OED recommendations.

3. Staffing and resource use

As mentioned earlier some evaluation work is undertaken in the regional, research, and policy departments of the Bank as part of their overall functions. The information below concerns solely the activities of OED.

OED's budget for professional staff for FY 1992 envisages 60.5 staff years (including 17.2 consultant years) for a total cost of $7.9 million. Included are 44 higher-level and 23 support staff positions. OED has also accepted support from other donor agencies for specific audits and studies of mutual interest, the largest of which has been the Canadian International Development Agency (CIDA). (Under an agreement which expired in August 1991, CIDA contributed about $870,000 to OED's work program over three years.)

In FY 1992, OED expects to review 296 PCRs and to audit 137 of them. In addition 28 studies will be prepared, 12 of which will be completed in FY 1992. Manpower use by type of report or study is as follows:

	In staff-weeks
Investment PCR review	0.9
Adjustment PCR review	1.2
Investment audit	5.6
Adjustment audit completed	11.3
Study	27.2

The above figures represent averages and the amount of staff time per review, audit, and especially study will vary according to the nature and complexity of the matter being studied. Among the studies under preparation, the Second SAL/SECAL Overview may be noted. The first was prepared in 1986 covering 15 of the early structural adjustment loans. The second report will be based on a broader sample of about 100 SALs/SECALs of which PCRs or audits have been completed. It will assess the quality of advice, the design and implementation of policy reforms, and the results achieved; analyze the shortcomings of adjustment programs and their causes; and offer lessons from experience for future SAL/SECAL operations.

4. An appraisal of Bank evaluation procedures

At an institutional level, there is little doubt that OED performs an important and valuable function, in enhancing the Bank's internal and external accountability and transparency (including in the eyes of credit markets where the Bank must seek its resources), in facilitating feedback from the lessons of past experience (and thus contributing to management), and in providing a systematic, quantified, and accessible documentary track record of the institution.

Nevertheless, appraising the effectiveness of the Bank's evaluation procedures is no easy task. The mass of documentary information can be perused, and discussions held with Bank staff who view evaluation from different standpoints, in particular the "producers" and the "consumers" of evaluation findings. But, there is no simple test of success and any conclusions based on a brief study such as this must remain largely judgmental, if not impressionistic.

There are, moreover, severe methodological obstacles to evaluation. Establishing causal links between Bank (and for that matter Fund) programs and economic performance is, as the economic literature shows, difficult enough in itself; seeking to relate any improvement in economic performance to better program design and implementation attributable to better evaluation compounds the difficulty. For instance, it is virtually impossible to establish whether, and if so to what extent, operations evaluation in the Bank has contributed to higher growth and investment in particular member countries, or has resulted in poverty alleviation.

A major problem is that evaluation cannot be based on a comparison of two static situations; one is dealing with a dynamic process. The picture is changing constantly; not only do circumstances and conditions evolve, but the Bank's concerns and priorities alter, member countries' needs change, and there are developments in OED's approach, initiatives, and personnel.

With these caveats in mind, this brief review of the Bank's evaluation procedures suggests a number of broad general conclusions.

First, the obvious point may be noted that the Bank has amassed an enormous corpus of experience in many fields and that this provides a rich fund from which to draw lessons for the future. In the view of some, this fund has not been sufficiently or profitably tapped. Operations evaluation, and in particular the activities of the OED, provides a convenient (though not perfect) way of employing and distilling the experience of past operations for more effective future projects and programs.

Second, the general consensus (within which there are variations) among the Bank staff is that operations evaluation as currently conducted is useful and contributes to improving future operations. Further, it is expected that the role of evaluation will increase in importance in view of the recent reorganization and decentralization of the Bank, 1/ as well as the evolution of criteria for judging Bank activities (for example, environmental issues).

Third, apart from the "bread and butter" functions of OED, that is the review and audit of PCRs, many departments use the OED as a channel of access to the experience of other sectors and regions; thus, OED is also an agent for generalizing within the operational departments the experience of Bank activities in a particular region or sector. Similarly, the cross-country and other studies undertaken by OED are useful and relevant to the work of the operational departments.

Finally, there is bound to be a degree of natural resistance on the part of operational staff to having their work evaluated; however, the unequivocal commitment of the Bank's management forcefully supported by the Board, to incorporating evaluation findings in operational work is key to the whole process. It promotes serious cooperation between the operational and evaluation staff; such cooperation appears to be candid, cordial, and productive.

On the other side of the ledger, Bank staff, both within OED and in the operational departments, do entertain a number of reservations, not all of which may relate to the activities of OED as such.

First, evaluations are based on ex post appraisals of projects and programs against original objectives, which may not have been clearly

---

1/ With the abolition last year of the Bank's Operations Department, there is no longer a centralized operational review body that surveys the quality of the Bank's portfolio and overall performance.

defined in the first place. That being so, the indicators of success also tend to be difficult to assess.

Second, there is a considerable time lag between implementation and the evaluation process. Typically, a project will be implemented over an eight-year period and an adjustment program over two to three years. The result is that the evaluation, especially in regard to the basic foundations of the program such as design, underlying assumptions regarding major economic magnitudes, and certain other aspects of the project, tend to be dated and to appear "old hat." Some question whether it is fair to point shortcomings of the program with the benefit of hindsight, and many years after the commencement of the project or program. Also, a need is felt for a more dynamic system of evaluation, one that takes into account changing circumstances and may, therefore, be more relevant and "actionable."

Third, while evaluation findings are now more targeted, issue-oriented, and focused, there is room for further progress. Moreover, in the view of some, such findings might be integrated more closely with the evaluation work that is undertaken by the research and policy departments of the Bank. Although this could serve to present a more united view of past experience, it could also dilute OED's independence.

Evaluation Practices of Other Institutions

In addition to the brief review of evaluation procedures in the World Bank, above, discussions were held with the directors and staff of evaluation units in a number of other multilateral and bilateral organizations, as well as with a cross-section of the "consumers" or intended beneficiaries of evaluation in these organizations. 1/ In several ways, the organizational and practical aspects of evaluation in these institutions differ from those of the World Bank. Some of the principal characteristics are summarized below.

1. Institutional features

Broadly speaking, all evaluation functions share the same basic *raison d'être*: to enhance the accountability; to learn from the past and thereby improve efficiency and effectiveness; to address special issues and problems; to provide a management tool; and to gauge the relevance, responsiveness and impact (direct and indirect) of the institution's activities, that is, to help determine whether the institution is serving its members and beneficiaries as it was intended to.

In general, evaluation units function outside the main operational framework of the institutions they serve. In all cases they enjoy a considerable degree of independence, if not in a formal sense, usually as regards their judgments, opinions, and findings. Equally important, while being outside the main lines of the work structure, evaluation units have full access to the information necessary to allow them to reach their determinations, including free dialogue with operational staff.

In terms of organizational placing, the set-up of the Asian Development Bank presents a number of interesting features. In contrast to the World Bank (where the Operations Evaluation Department reports directly to the Executive Board), the Post Evaluation Office of the ADB reports to management, i.e., to the President of the ADB. PEO's evaluation reports, studies, and findings are reviewed and discussed in formal meetings of the committee of ADB's executive board, the Audit Committee. This Committee also discusses PEO's annual work program and may make suggestions for future studies. Although the PEO reports to the President, who forwards the reports to the Committee, by tradition he does not modify the findings--or criticisms--of the PEO. The full executive board discusses the annual report of the PEO.

1/ The organizations included the Asian Development Bank (ADB), the Inter-American Development Bank (IDB), the African Development Bank (AfDB), the OECD, the United States Agency for International Development (AID), the Overseas Development Administration of the United Kingdom (ODA), the Swedish International Development Authority (SIDA), and the Japanese Overseas Economic Cooperation Fund (OECF).

Although the PEO is an integral part of the ADB's staff structure, the members of the Audit Committee appear satisfied with the judgmental independence of the PEO. The Committee and the PEO have cordial and candid working relations. This, together with the strong support of the ADB's management for the evaluation function, has served to make the current arrangements acceptable to all sides. Indeed, some consider that, for a relatively small institution such as the ADB, these arrangements have certain advantages in that an evaluation unit completely outside the staff structure (and reporting to the executive board) could run the risk of becoming isolated and ineffective. 1/

2. Working practices

The emphasis in all cases is on evaluation of performance, in particular, of projects but also, increasingly, of sector programs. In a few cases other activities (such as technical assistance) have also been evaluated. Where broader evaluations are made, such as cross-country studies, experience of particular sectors, or assessments of general policies and procedures, these are often based on individual project or program evaluations as their basic element. These evaluations, in turn, employ the self-evaluation of operational departments (that is the project completion report or PCR) as their point of departure.

In practice, no institution or organization evaluates its entire range of operational activities; a selection has to be made. In the ADB, for instance, which evaluates about 50 percent of all projects, the selection is random. In other cases, projects may be evaluated if they encounter significant problems, or if a change of approach is desired. In some aid agencies, operational departments themselves may suggest particular projects as suitable candidates for evaluation.

In method and analysis, the approach differs from purely descriptive accounts of performance at one end, to specification of results in terms of qualitative indicators, at the other. Often, more than one methodology is used. Experience seems to suggest that the more specific the project or program being evaluated, the more objective and quantitative the evaluation can be; as a corollary, the broader the operation, the more judgmental--and often difficult--the evaluation. Thus, evaluating a fertilizer project is more straightforward than a program to reduce subsidies in the agricultural sector, which in turn is easier than assessing a macroeconomic stabilization program.

The most readily usable and understood method of evaluation is to compare the original objectives of the project or program with the actual

---

1/ The Operation Evaluation Office of the IDB is also an integral part of its staff structure, reporting through the Controller to management. But the IDB has the unique feature of having two evaluation units; the Office of External Review and Evaluation reports directly to the executive board.

outcome. Under this broad umbrella, a large number of specific aspects can be assessed, including, for instance, cost overruns, delays in execution, design of the project, implementation, and cost effectiveness. The economic rate of return for a particular project is often used as a quantitative indicator of success or failure. For instance, in the ADB a rate of return of 8 percent, provided there are some unquantifiable benefits and reasonable prospects for achieving them, qualifies as success; a rate of between 4 and 8 percent partial success; and a rate below 4 as failure. Using the economic rate of return, however, has its limitations; for instance, it is not appropriate in a health project.

Another, complementary, indicator is project or program impact. This concept can be applied to a comparison of the actual impact with the intended impact, but may also be expanded to take into account broader socio-economic goals. Typically, impact studies are undertaken some time after the completion of the project. Similarly, the concept of sustainability is one that is sometimes used in assessing the longer-term effectiveness of a particular project or program.

Other approaches (such as "before/after" comparisons, the "logical-framework approach," and counterfactual analysis) are sometimes used. In several evaluation units, the methodology of evaluation itself is the subject of study and investigation.

As mentioned above, evaluation units enjoy free access to all available information and, in their evaluative work, send missions to the field for discussions with the authorities and intended beneficiaries of projects. In the multilateral agencies the work of evaluation is undertaken by the staff of the units concerned, although some use is made of outside consultants. Aid agencies, on the other hand, make extensive use of consultants who often sign their reports. At the Swedish International Development Authority, in fact, the operational departments may commission their own evaluation reports; in such instances it is the task of the Evaluation Department to ensure that evaluation is carried out, that its quality is acceptable, and that the terms of reference for consultants are appropriate.

3. Dissemination and absorption

The importance of presenting and disseminating evaluation findings, and ensuring that they are taken into account in future operations, is keenly appreciated in all organizations. Generally, evaluation reports on particular projects, broader sectoral or policy studies, and reports that summarize evaluative activities on an annual basis, are distributed to the board, management, and senior staff of the organization as well as to the officials of the borrowing country. Some of these reports, especially general studies or annual reports of findings, are published; the country-specific evaluations generally are not but are available to interested parties.

In addition some evaluation units seek to facilitate access to their findings by preparing readable summaries and précis and by making findings available on computer networks. Workshops, seminars, and other discussion forums also play a part in disseminating evaluation findings, as do informal individual contacts among the staff. On a more formal level, evaluations findings may be discussed by an oversight or governing body of the evaluation unit. In the case of the ADB, for instance, individual evaluation reports are discussed by the Audit Committee of the executive board where the evaluation staff can defend their findings in the presence of the operational staff whose project or program is the subject of evaluation) while the full executive board discusses the annual report of the evaluation unit. Also, a standing committee, chaired by the President and attended by senior staff, takes up the broader issues that have been studied by the evaluation unit.

The foregoing procedures are, in some organizations, supplemented by specific arrangements to ensure that evaluation findings are taken into account by the operational staff and that they are reflected in proposals for new projects or programs. However, practices vary considerably; in some organizations there is no formal machinery for this type of feedback.

In contrast to the practice of the World Bank, both the ADB and the IDB have adopted procedures that allow evaluation units to have an input at the preparative and pre-appraisal stages of new projects and programs. In the case of the ADB, the Post Evaluation Office reviews and comments on papers pertaining to new projects at several stages prior to approval, seeking to ensure that past findings have been incorporated. In the IDB, the Operation Evaluation Office's findings are systematically fed into the project preparation, approval, and execution cycle (e.g., at meetings of the loan committee and programming committee). Also, in both institutions a periodic status of implementation of executive board (or board committee) recommendations is prepared (by the office of the Controller in the IDB and the Secretary's office at the ADB).

4. Reception by operational staff

Individual views among the operational staff concerning the benefits of evaluation findings vary considerably: some welcome evaluation, others are resistant and defensive about it, and, in many cases, staff confess to being originally skeptic but having become converts.

The general consensus would seem to be that an independent and systematic way of making operational staff aware of the lessons of the past is useful and can lead to improved program design and implementation, that in some instances it can help uncover deficiencies and pinpoint problems, and that, in a general way, it helps keep operational staff "on their toes" even if critical evaluations are not a source of comfort. At the same time, there are reservations about some evaluation findings (in particular, that they are often dated, uninformed and perhaps simplistic). Findings bearing on broader and longer-terms issues, including sectoral summaries and studies that provide a deeper understanding of success or failure of programs and projects, are found to be of greater interest, while (as in the case of the World Bank) the lack of sufficient time on the part of operational staff to adequately absorb the output of the evaluation units is noted by many. This highlights the importance of effective methods of dissemination.

In almost all cases, the evaluation function was introduced not at the inception of the institution or agency in question, but at a later date. It is, however, to be noted that evaluation, as an integral part of the ways in which the institution carries out its mandate and serves its beneficiaries, is an accepted fact, even if it had been originally resisted.